future internet

Special Issue Reprint

Artificial Intelligence and Control Systems for Industry 4.0 and 5.0

Edited by
Filipe Pereira and Paulo Leitao

mdpi.com/journal/futureinternet

MDPI

Artificial Intelligence and Control Systems for Industry 4.0 and 5.0

Artificial Intelligence and Control Systems for Industry 4.0 and 5.0

Guest Editors

Filipe Pereira
Paulo Leitao

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Filipe Pereira
Faculdade de Engenharia
Universidade do Porto (FEUP)
Porto
Portugal

Paulo Leitao
Research Center in
Digitalization and Intelligent
Robotics (CeDRI)
Instituto Politécnico de Bragança
Bragança
Portugal

Editorial Office

MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Future Internet* (ISSN 1999-5903), freely accessible at: https://www.mdpi.com/journal/futureinternet/special_issues/0FVUL7IS43.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-8869-6 (Hbk)
ISBN 978-3-7258-8870-2 (PDF)
https://doi.org/10.3390/books978-3-7258-8870-2

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editors . . . . . **vii**

Preface . . . . . **ix**

Filipe Pereira and Paulo Leitão
Artificial Intelligence and Control Systems for Industry 4.0 and 5.0: Recent Advances, Knowledge Gaps, and Future Research Directions
Reprinted from: *Future Internet* **2026**, *18*, 380, https://doi.org/10.3390/fi18070380 . . . . . **1**

Riccardo Bacca, Andrea Melis, Lorenzo Rinieri, Roberto Girau, Marco Prandini, Franco Callegati
Toward Secure Software-Defined Industrial Networks Through Asset Administration Shell Digital Twins
Reprinted from: *Future Internet* **2026**, *18*, 347, https://doi.org/10.3390/fi18070347 . . . . . **7**

Vaclav Mach, Jiri Vojtesek, Milan Adamek, Pavel Drabek, Pavel Stoklasek, Stepan Dlabaja, et al.
Smart IoT-Based Temperature-Sensing Device for Energy-Efficient Glass Window Monitoring
Reprinted from: *Future Internet* **2025**, *17*, 576, https://doi.org/10.3390/fi17120576 . . . . . **38**

Ádám Francuz and Tamás Bányai
Intelligent Control Approaches for Warehouse Performance Optimisation in Industry 4.0 Using Machine Learning
Reprinted from: *Future Internet* **2025**, *17*, 468, https://doi.org/10.3390/fi17100468 . . . . . **57**

Andrés Fernández-Miguel, Susana Ortíz-Marcos, Mariano Jiménez-Calzado, Alfonso P. Fernández del Hoyo, Fernando E. García-Muiña and Davide Settembre-Blundo
Toward the Theoretical Foundations of Industry 6.0: A Framework for AI-Driven Decentralized Manufacturing Control
Reprinted from: *Future Internet* **2025**, *17*, 455, https://doi.org/10.3390/fi17100455 . . . . . **80**

Olha Kozina, José Machado, Maksym Volk, Hennadii Heiko, Volodymyr Panchenko, Mykyta Kozin and Maryna Ivanova
Opportunities for Adapting Data Write Latency in Geo-Distributed Replicas of Multicloud Systems
Reprinted from: *Future Internet* **2025**, *17*, 442, https://doi.org/10.3390/fi17100442 . . . . . **107**

Henrique C. Ferreira and Ramiro S. Barbosa
Deep Reinforcement Learning for Adaptive Robotic Grasping and Post-Grasp Manipulation in Simulated Dynamic Environments
Reprinted from: *Future Internet* **2025**, *17*, 437, https://doi.org/10.3390/fi17100437 . . . . . **134**

Andrés Fernández-Miguel, Fernando Enrique García-Muiña, Susana Ortíz-Marcos, Mariano Jiménez-Calzado, Alfonso P. Fernández del Hoyo and Davide Settembre-Blundo
AI-Driven Transformations in Manufacturing: Bridging Industry 4.0, 5.0, and 6.0 in Sustainable Value Chains
Reprinted from: *Future Internet* **2025**, *17*, 430, https://doi.org/10.3390/fi17090430 . . . . . **152**

Lucas Sakurada, Fernando De la Prieta and Paulo Leitao
The Role of Multi-Agent Systems in Realizing Asset Administration Shell Type 3
Reprinted from: *Future Internet* **2025**, *17*, 270, https://doi.org/10.3390/fi17070270 . . . . . **174**

Ioana-Livia Stefan, Andrei Mateescu, Ionut Lentoiu, Silviu Raileanu, Florin Daniel Anton, Dragos Constantin Popescu and Ioan Stefan Sacala
Implementation and Performance Analysis of an Industrial Robot's Vision System Based on Cloud Vision Services
Reprinted from: *Future Internet* **2025**, *17*, 200, https://doi.org/10.3390/fi17050200 . . . . . **200**

About the Editors

Filipe Pereira

Filipe Pereira is a professor and researcher at the Faculty of Engineering of the University of Porto and industrial consultant in Electrical and Computer Engineering with extensive experience in industrial automation, robotics, artificial intelligence, industrial communications, control systems, Industrial Internet of Things, computer vision, and smart manufacturing. His research focuses on Industry 4.0 and Industry 5.0 technologies, including machine learning, deep learning, collaborative robotics, digital twins, intelligent inspection systems, and cyber-physical production systems. He has participated in numerous industrial R&D projects developed in collaboration with companies and research institutions and has authored scientific publications in international journals and conferences. His current research interests include intelligent manufacturing systems, AI-enabled industrial automation, industrial digitalization, and advanced robotics.

Paulo Leitao

Paulo Leitao is a professor and researcher at the Polytechnic Institute of Bragança, Portugal. His research interests include industrial automation, industrial informatics, distributed intelligent systems, cyber-physical production systems, Industry 4.0, multi-agent systems, Industrial Internet of Things, and digital transformation of manufacturing systems. He has coordinated and participated in numerous national and European R&D projects and has authored a significant number of scientific publications in leading international journals and conferences. His work focuses on intelligent manufacturing, interoperability, industrial communication systems, and advanced industrial control architectures.

Preface

The rapid evolution of Artificial Intelligence has significantly accelerated the digital transformation of industrial systems. Technologies such as machine learning, deep learning, computer vision, collaborative robotics, Industrial Internet of Things, cloud computing, digital twins, and advanced control architectures are redefining manufacturing paradigms and enabling increasingly autonomous, resilient, sustainable, and human-centric production environments.

This Reprint brings together high-quality research contributions that explore recent advances in artificial intelligence and intelligent control systems for Industry 4.0 and the emerging Industry 5.0 paradigm. The published works address theoretical developments, innovative methodologies, practical implementations, and experimental validations across multiple industrial domains, providing a comprehensive overview of current scientific progress.

Beyond presenting state-of-the-art solutions, the contributions identify important research challenges related to interoperability, cybersecurity, trustworthy AI, explainability, sustainability, edge intelligence, digital transformation, and human–machine collaboration. These topics are expected to play an increasingly important role in the next generation of intelligent manufacturing systems.

We hope that this Reprint will provide researchers, engineers, practitioners, and graduate students with valuable insights into current developments while simultaneously inspiring further interdisciplinary research and industrial innovation. We sincerely thank all authors, reviewers, and the editorial staff of *Future Internet* for their dedication and valuable contributions throughout the publication process.

Filipe Pereira and Paulo Leitao
Guest Editors

future internet

MDPI

Editorial

Artificial Intelligence and Control Systems for Industry 4.0 and 5.0: Recent Advances, Knowledge Gaps, and Future Research Directions

Filipe Pereira [1,2,3,4,*] and Paulo Leitão [4,5]

[1] Faculdade de Engenharia, Universidade do Porto (FEUP), Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[2] Instituto Superior de Engenharia do Porto (ISEP), Polytechnic of Porto, Rua Dr. António Bernardino de Almeida, 4249-015 Porto, Portugal
[3] Institute of Science and Innovation in Mechanical and Industrial Engineering (INEGI - LAETA), Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[4] Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal; pleitao@ipb.pt
[5] Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
* Correspondence: filipepereira@fe.up.pt

1. Introduction

Artificial intelligence, control systems, industrial IoT, digital twins, and cyber–physical production systems are reshaping the technological foundations of advanced manufacturing. This Editorial synthesizes the contributions published in the Special Issue "Artificial Intelligence and Control Systems for Industry 4.0 and 5.0" and positions them within the transition from highly connected Industry 4.0 environments to more human-centric, sustainable, and resilient Industry 5.0 systems. The papers in this Special Issue address cloud-based robotic vision, multi-agent systems and Asset Administration Shells, AI-driven value chains, deep reinforcement learning for robotic manipulation, multi-cloud data latency, decentralized manufacturing control, machine-learning-based warehouse optimization, smart IoT sensing for energy efficiency, and secure software-defined industrial networks. Together, these contributions highlight both the maturity of AI-enabled industrial applications and the remaining challenges associated with trustworthiness, interoperability, cybersecurity, scalability, real-time control, and socio-technical integration. Future research should therefore prioritize verifiable AI for closed-loop control, standardized digital-twin infrastructures, edge-cloud orchestration, human-in-the-loop autonomy, cyber-resilient architectures, and sustainability-aware industrial intelligence.

Industry 4.0 established a paradigm in which manufacturing systems are increasingly connected, data-driven, and integrated through cyber–physical systems, industrial IoT, cloud services, automation technologies, and advanced analytics. However, the transition to Industry 5.0 has widened the scope of industrial digitalization. Rather than focusing only on connectivity and efficiency, current industrial research increasingly emphasizes human-centricity, sustainability, resilience, trustworthy autonomy, and the capacity of production systems to adapt to disturbances while preserving social and environmental value [1].

In this context, artificial intelligence and control systems are not independent technological domains. Their convergence is essential for the design of adaptive, autonomous, and collaborative industrial systems. AI provides perception, prediction, diagnosis, optimization, and decision-support functions, whereas control systems translate these functions into

safe, timely, and reliable action in physical production environments. The integration of both domains is therefore central to the next generation of industrial robots, smart factories, digital twins, distributed manufacturing systems, energy-aware infrastructures, and secure industrial networks.

The Special Issue "Artificial Intelligence and Control Systems for Industry 4.0 and 5.0" was launched to collect recent advances at this intersection. The nine contributions published in the Special Issue demonstrate that the field is progressing from conceptual digitalization to deployable architectures, algorithms, and industrial use cases. At the same time, they show that major scientific and engineering challenges remain open, particularly when AI is embedded in cyber–physical systems that require real-time performance, explainability, interoperability, safety, cybersecurity, and long-term maintainability.

2. Contributions in This Special Issue

The contributions in this Special Issue can be grouped into four complementary research streams: AI-enabled perception and robotics, digital twins and decentralized industrial intelligence, data-driven optimization and industrial infrastructures, and sustainability- and security-oriented Industry 5.0 systems.

Table 1 summarizes the main themes and scientific contributions of the papers published in this Special Issue. These contributions can be grouped into four complementary research streams: AI-enabled perception and robotics, digital twins and decentralized industrial intelligence, data-driven optimization and industrial infrastructures, and sustainability- and security-oriented Industry 5.0 systems.

Table 1. Overview of the main themes and contributions of the papers published in the Special Issue "Artificial Intelligence and Control Systems for Industry 4.0 and 5.0".

Reference	Main Theme	Contribution to the Special Issue
[2]	Cloud-based robotic vision	Assesses cloud vision services for industrial robot perception and object sorting, linking AI perception with industrial robotic execution.
[3]	Multi-agent systems and AAS	Discusses the role of multi-agent systems in creating active Asset Administration Shell Type 3 entities for more autonomous industrial digital twins.
[4]	AI-driven sustainable value chains	Examines how AI, digital twins, and cognitive automation can support predictive maintenance, circularity, and resilient value-chain transformation.
[5]	Deep reinforcement learning for robotics	Investigates adaptive robotic grasping and post-grasp manipulation in simulated dynamic environments, contributing to autonomous manipulation research.
[6]	Multi-cloud industrial data infrastructures	Addresses write-latency adaptation in geo-distributed multi-cloud replicas, a relevant problem for resilient industrial data management.
[7]	AI-driven decentralized manufacturing control	Develops theoretical foundations for decentralized AI-enabled manufacturing control and extends the discussion to future Industry 6.0 frameworks.
[8]	Machine learning for warehouse optimization	Uses machine learning to model and optimize warehouse performance, demonstrating interpretable data-driven decision support for logistics.
[9]	Smart IoT sensing for energy efficiency	Presents an IoT-based temperature-sensing device for glass-window monitoring, connecting industrial sensing, energy efficiency, and smart-building operation.
[10]	Secure software-defined industrial networks	Combines software-defined networking and AAS digital twins to address cybersecurity and resilience in industrial network infrastructures.

https://doi.org/10.3390/fi18070380

The robotics-oriented papers demonstrate the increasing importance of AI regarding perception and manipulation. Cloud-based vision services can reduce the barrier to implementing advanced visual recognition in robotic cells, but their deployment also raises questions about latency, robustness, data governance, and dependency on external infrastructures [2]. In parallel, deep reinforcement learning provides a route toward more adaptive robotic grasping and manipulation, especially in dynamic or uncertain environments, but still requires careful validation before transfer from simulation to production [5].

This direction is also consistent with recent work integrating trained deep learning vision systems with collaborative robots for real-time sorting and quality inspection [11]. Such contributions reinforce the need to evaluate AI-based perception not only through model-level metrics but also through deployment-oriented indicators such as cycle time, reliability, integration with robotic motion, and operational robustness.

The second group of papers addresses decentralized intelligence, digital twins, and industrial interoperability. The Asset Administration Shell is a key element for standardized industrial digital representations, but active and autonomous AAS implementations require mechanisms for negotiation, coordination, and decision-making. Multi-agent systems are therefore highly relevant for implementing AAS Type 3 functionalities [3]. This perspective is further extended by work on AI-driven decentralized manufacturing control and the use of AAS digital twins in secure software-defined industrial networks [7,10].

The third stream concerns data-driven industrial optimization and the infrastructures needed to support it. Machine learning can transform warehouse operation from rule-based management into predictive and optimization-oriented decision support [8]. However, such approaches depend on data infrastructures that can guarantee performance, consistency, and availability across distributed environments. The contribution on geo-distributed multi-cloud replicas therefore highlights an important enabling layer for future industrial platforms [6].

Finally, the Special Issue connects AI and control systems with sustainability and Industry 5.0 priorities. AI-driven transformations in manufacturing value chains can support predictive maintenance, resource efficiency, circular economy practices, and regulatory alignment [4]. Similarly, smart IoT sensing for energy-efficient monitoring shows how low-level sensing infrastructures can contribute to broader energy-management and sustainability objectives [9].

3. Knowledge Gaps and Open Challenges

Taken together, the papers indicate that AI and control systems are already producing practical results for industrial environments. Nevertheless, several gaps remain insufficiently sealed.

- **Trustworthy AI for closed-loop systems:** Industrial AI is increasingly used for perception, prediction, and optimization, but its integration into closed-loop control requires stronger evidence of robustness, explainability, safety, and performance under distribution shifts.
- **Interoperability between digital twins, AAS, and legacy systems:** Digital twins and AAS-based architectures are promising, but their implementation remains fragmented across platforms, communication protocols, semantic models, and organizational maturity levels.
- **Real-time edge-cloud and multi-cloud orchestration:** Industrial applications often require low latency, deterministic behavior, and high availability. These requirements are difficult to reconcile with cloud-based AI services and geographically distributed data infrastructures.

https://doi.org/10.3390/fi18070380

- **Cybersecurity and resilience in convergent OT/IT environments:** As industrial networks become software-defined, connected, and AI-enabled, the attack surface expands. Security must be designed as a system property rather than added after deployment.
- **Human-centricity beyond conceptual statements:** Industry 5.0 requires measurable integration of human factors, operator trust, ergonomics, skills, collaborative autonomy, and ethical decision-making. These dimensions are still less mature than purely technical optimization metrics. The educational and skills dimension is particularly relevant because AI-supported learning environments for electropneumatic control systems can help prepare engineers to design, diagnose, and maintain intelligent automation systems [12].
- **Sustainability-aware control and optimization:** Energy efficiency, circularity, and environmental impacts should become explicit objectives in AI-enabled industrial control rather than secondary evaluation criteria.
- **Reproducibility and benchmarking:** Many industrial AI studies remain difficult to reproduce because datasets, realistic benchmarks, deployment constraints, and long-term industrial validation protocols are limited.

4. Future Research Directions

Future research on AI and control systems for Industry 4.0 and 5.0 should move from isolated demonstrations to validated, interoperable, and human-compatible industrial ecosystems. The following priorities are particularly relevant.

- **Verifiable and explainable AI-enabled control:** New methods are required to combine machine learning with model-based control, formal verification, uncertainty quantification, and runtime monitoring. This is especially important for safety-critical robotic cells, adaptive production systems, and autonomous logistics.
- **Operational digital twins and AAS-based autonomy:** Digital twins should evolve from monitoring dashboards into operational entities capable of supporting diagnosis, negotiation, optimization, and autonomous or semi-autonomous control. Multi-agent systems can provide an architectural basis for this transition.
- **Edge-fog-cloud intelligence for industrial systems:** The next generation of industrial AI should distribute computation across sensors, controllers, edge nodes, fog layers, and cloud services according to latency, privacy, reliability, and cost constraints.
- **Cyber-resilient industrial networks:** Future industrial control infrastructures should integrate software-defined networking, zero-trust principles, attack modeling, anomaly detection, and digital-twin-based security assessment.
- **Human-in-the-loop and human-on-the-loop autonomy:** Industry 5.0 systems must support operators as active participants in decision-making. Research should address adaptive interfaces, explainable recommendations, skill augmentation, collaborative robotics, and measurable human-well-being outcomes.
- **Sustainability and circularity as control objectives:** AI-driven optimization should explicitly include energy consumption, emissions, material reuse, maintenance strategies, and circular economy indicators. This requires multi-objective control frameworks and life-cycle-aware decision support.
- **Industrial foundation models and agentic AI with safeguards:** Large AI models and autonomous software agents may support planning, troubleshooting, documentation, coding, and operator assistance, but their deployment in industrial environments requires domain grounding, verification, access control, and fail-safe integration.

https://doi.org/10.3390/fi18070380

- **Open benchmarks, datasets, and deployment evidence:** The field would benefit from shared datasets, simulation-to-real benchmarks, reference architectures, and longitudinal industrial case studies that report not only algorithmic accuracy but also downtime reduction, energy savings, productivity, safety, and maintenance impacts.

5. Conclusions

The Special Issue "Artificial Intelligence and Control Systems for Industry 4.0 and 5.0" reflects a field in rapid transition. AI is no longer limited to offline analytics; it is becoming embedded in robotic perception, industrial decision support, digital twins, control architectures, logistics systems, sensing devices, and cybersecurity mechanisms. At the same time, control systems are evolving from deterministic automation layers into adaptive and distributed infrastructures capable of interacting with data-driven intelligence.

The papers collected in this Special Issue demonstrate the breadth of this transformation. They also clarify that the success of AI-enabled industrial systems will depend not only on algorithmic performance but also on trustworthy deployment, interoperability, real-time behavior, cyber-resilience, sustainability, and meaningful human integration. These challenges define the research agenda for the next generation of Industry 4.0 and 5.0 systems and provide a strong foundation for future work at the intersection of artificial intelligence, control engineering, and industrial digital transformation.

Author Contributions: Conceptualization, F.P. and P.L.; writing—original draft, F.P.; writing—review and editing, F.P. and P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Breque, M.; De Nul, L.; Petridis, A. *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*; European Commission, Directorate-General for Research and Innovation: Luxembourg, 2021. [CrossRef] [PubMed]
2. Stefan, I.-L.; Mateescu, A.; Lentoiu, I.; Raileanu, S.; Anton, F.D.; Popescu, D.C.; Sacala, I.S. Implementation and Performance Analysis of an Industrial Robot's Vision System Based on Cloud Vision Services. *Future Internet* **2025**, *17*, 200. [CrossRef]
3. Sakurada, L.; De la Prieta, F.; Leitão, P. The Role of Multi-Agent Systems in Realizing Asset Administration Shell Type 3. *Future Internet* **2025**, *17*, 270. [CrossRef]
4. Fernández-Miguel, A.; García-Muiña, F.E.; Ortíz-Marcos, S.; Jiménez-Calzado, M.; Fernández del Hoyo, A.P.; Settembre-Blundo, D. AI-Driven Transformations in Manufacturing: Bridging Industry 4.0, 5.0, and 6.0 in Sustainable Value Chains. *Future Internet* **2025**, *17*, 430. [CrossRef]
5. Ferreira, H.C.; Barbosa, R.S. Deep Reinforcement Learning for Adaptive Robotic Grasping and Post-Grasp Manipulation in Simulated Dynamic Environments. *Future Internet* **2025**, *17*, 437. [CrossRef]
6. Kozina, O.; Machado, J.; Volk, M.; Heiko, H.; Panchenko, V.; Kozin, M.; Ivanova, M. Opportunities for Adapting Data Write Latency in Geo-Distributed Replicas of Multicloud Systems. *Future Internet* **2025**, *17*, 442. [CrossRef]
7. Fernández-Miguel, A.; Ortíz-Marcos, S.; Jiménez-Calzado, M.; Fernández del Hoyo, A.P.; García-Muiña, F.E.; Settembre-Blundo, D. Toward the Theoretical Foundations of Industry 6.0: A Framework for AI-Driven Decentralized Manufacturing Control. *Future Internet* **2025**, *17*, 455. [CrossRef]
8. Francuz, Á.; Bányai, T. Intelligent Control Approaches for Warehouse Performance Optimisation in Industry 4.0 Using Machine Learning. *Future Internet* **2025**, *17*, 468. [CrossRef]
9. Mach, V.; Vojtesek, J.; Adamek, M.; Drabek, P.; Stoklasek, P.; Dlabaja, S.; Kopecek, L.; Mizera, A. Smart IoT-Based Temperature-Sensing Device for Energy-Efficient Glass Window Monitoring. *Future Internet* **2025**, *17*, 576. [CrossRef]
10. Bacca, R.; Melis, A.; Rinieri, L.; Girau, R.; Prandini, M.; Callegati, F. Toward Secure Software-Defined Industrial Networks Through Asset Administration Shell Digital Twins. *Future Internet* **2026**, *18*, 347. [CrossRef]

11. Terras, N.; Pereira, F.; Ramos Silva, A.; Santos, A.A.; Lopes, A.M.; Silva, A.F.d.; Cartal, L.A.; Apostolescu, T.C.; Badea, F.; Machado, J. Integration of Deep Learning Vision Systems in Collaborative Robotics for Real-Time Applications. *Appl. Sci.* **2025**, *15*, 1336. [CrossRef]
12. Silva, A.F.; Santos, A.A.; Pereira, F.; Felgueiras, C.; Moreira, A.P. Application of Artificial Intelligence to Improve Skills in Electropneumatic Control Systems to Engineering Students. In *Proceedings of the 3rd International Conference on Innovation in Engineering (ICIE 2024), Povoação, Portugal, 26–28 June 2024*; Lecture Notes in Mechanical Engineering; Springer: Cham, Switzerland, 2024. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

Toward Secure Software-Defined Industrial Networks Through Asset Administration Shell Digital Twins

Riccardo Bacca [1], Andrea Melis [2], Lorenzo Rinieri [2,*], Roberto Girau [2], Marco Prandini [2] and Franco Callegati [2]

[1] Centro Interdipartimentale di Ricerca Industriale—CIRI ICT, Alma Mater Studiorum—Università di Bologna, 40126 Bologna, Italy; riccardo.bacca@unibo.it
[2] Department of Computer Science and Engineering (DISI), Alma Mater Studiorum—Università di Bologna, 40126 Bologna, Italy; a.melis@unibo.it (A.M.); roberto.girau@unibo.it (R.G.); marco.prandini@unibo.it (M.P.); franco.callegati@unibo.it (F.C.)
* Correspondence: lorenzo.rinieri@unibo.it

Abstract

Industrial digitalization is moving from Industry 4.0 toward Industry 5.0's emphasis on resilience, human-centric operation, and sustainability. This shift is enabled by the convergence of Operational Technology and Information Technology, but this integration also broadens the exposure of industrial infrastructures to cyber threats targeting communication integrity and process continuity. Mitigating these risks requires network control that is both programmable and aware of each asset's operational context. However, there is still a lack of operational interfaces that translate the semantics of industrial assets into programmable, runtime-enforceable network behavior. In this paper, following a Design Science Research methodology, we introduce an asset-aware, closed-loop network control abstraction in which the industrial network itself is modeled as a managed asset through Asset Administration Shells. Asset state, lifecycle phase, and operational intent are translated into network policies enforced at runtime on programmable data planes, while in-network telemetry is exposed at the asset level and correlated with operational metrics. We validate the abstraction on a hybrid testbed that combines virtualized components with industrial-grade hardware and virtualized 5G connectivity, through three security-oriented use cases: (i) asset-driven customization of forwarding policies; (ii) human-centric secure maintenance with controlled remote access over 5G; and (iii) anomaly detection and isolation based on cross-layer telemetry correlation. The results show that asset-level operations can drive programmable network enforcement and make network telemetry available at the asset layer. Finally, the work outlines a first step toward standardizing network-oriented asset submodels by separating control-plane operations from data-plane state and telemetry.

Keywords: industrial security; P4; software-defined networking; asset administration shell; digital twin; Industry 4.0; Industry 5.0; design science research

1. Introduction

Operational Technology (OT) networks interconnect industrial control systems, machines, sensors, and actuators, whereas Information Technology (IT) systems provide enterprise services, data analytics, supervision, and remote access capabilities [1,2]. Historically, OT networks were organized as isolated or weakly interconnected production cells. This organization reduced cyber exposure, but it also limited unified asset visibility,

dynamic network management, and semantic interoperability across industrial systems. Industry 4.0 has changed this model by requiring continuous data exchange within OT environments and across the OT/IT boundary, enabling predictive maintenance, flexible production, and data-driven decision-making [3].

This convergence is essential, but it introduces two distinct challenges. The first concerns OT network management: industrial networks still suffer from limited unified visibility over assets, proprietary or non-standardized endpoint discovery, static configuration practices, and poor telemetry integration. These limitations make it difficult to adapt network behavior to the actual state and requirements of the production environment. The second concerns IT/OT interconnection: as OT systems become reachable from IT services, edge platforms, and cloud services, they are increasingly exposed to cyber threats, unauthorized access, and disruptions that can compromise production continuity [4]. Addressing these challenges requires network-control mechanisms that are both programmable and aware of the operational context of industrial assets.

A promising basis for asset awareness is the Asset Administration Shell (AAS), introduced in the Reference Architecture Model for Industrie 4.0 (RAMI 4.0) as the digital representation of an asset that becomes an Industrie 4.0 component [5]. The AAS exposes the data, functions, and relationships that characterize an asset and its interactions with other assets in a system or process [6,7]. In parallel, Software-Defined Networking (SDN) and programmable data planes provide runtime control over forwarding behavior, enabling network policies to be installed, modified, and monitored through software interfaces [8,9].

Despite these complementary capabilities, a gap remains between industrial asset semantics and programmable network enforcement. Existing AAS-based approaches generally represent machines, services, or lifecycle information, but they do not usually model the communication network itself as an AAS-managed industrial asset. Conversely, SDN and P4 provide powerful enforcement mechanisms, but their policies are typically expressed through network-level identifiers such as addresses, ports, tables, and forwarding rules, rather than through standardized industrial asset semantics. As a result, there is no commonly adopted AAS model for programmable network control that separates control-plane operations from data-plane state and telemetry.

The research question addressed in this paper is therefore: how can standardized asset semantics be coupled with programmable networking to support closed-loop, asset-aware security control in converged OT/IT environments? We address this question using a Design Science Research methodology, since the contribution is an engineered artifact that is designed, implemented, and evaluated in a representative technical environment [10,11].

The artifact presented in this work is an AAS-driven network control abstraction in which the OT network is represented as an industrial asset with dedicated AASs for control-plane and data-plane functions. AAS operations are translated into SDN/P4 enforcement actions, while network telemetry is exposed back through the AAS layer and correlated with asset-side operational metrics. This creates a bidirectional coupling between industrial asset semantics and programmable network behavior.

The contribution is positioned as an Industry 4.5 step: it validates the semantic and programmable mechanisms required by Industry 4.0 while providing a technical basis for selected Industry 5.0 requirements, including resilience, human-centric operation, and sustainability-aware network control. The paper makes four specific contributions:

1. It defines a network AAS abstraction that binds asset-level information to programmable network policies;
2. It implements the abstraction using Eclipse BaSyx, P4Runtime, Kathará, and virtualized 5G components;

https://doi.org/10.3390/fi18070347

3. It validates the framework through three security-oriented use cases, including hardware-in-the-loop validation with an industrial Revolution Pi device;
4. It identifies a first step toward standardizing network-oriented AAS submodels by separating control-plane operations from data-plane state and telemetry.

The remainder of this paper is organized as follows. Section 2 reviews the related work and highlights the research gaps addressed by the study. Section 3 presents the problem statement, methodology, and threat model. Section 4 describes the architecture. Section 5 presents the testbed, and Section 6 reports the validation use cases and results. Section 7 discusses limitations, maturity, Industry 5.0 implications, and standardization opportunities. Section 8 concludes the paper.

2. Related Work

The related work is organized around the functional chain required by the proposed artifact. First, industrial assets must be represented through a standardized semantic interface. Second, the network must be programmable at runtime. Third, monitoring information must be correlated across asset and network layers. Fourth, remote maintenance and 5G connectivity introduce new access-control requirements. Finally, Industry 5.0 extends the evaluation criteria from automation alone toward resilience, human-centricity, and sustainability. This structure allows us to identify the specific gap addressed in this paper: the lack of an operational binding between AAS-based asset semantics and programmable network enforcement.

In the Industry 4.0 stack, the Asset Administration Shell is the standardized digital counterpart of a physical or logical asset, exposing semantically structured information and services across the asset lifecycle [12,13]. AAS research has addressed interoperability, lifecycle data, safety, and security descriptors [14,15], and the AAS structure has been standardized in IEC 63278-1 [16]. These works establish the AAS as an industrial digital-twin interface, but they do not generally use AAS operations as runtime inputs for network policy enforcement.

Programmable networking provides the enforcement substrate needed to act on such asset-level information. SDN separates control logic from forwarding behavior, while P4 enables programmable data-plane functions such as filtering, counters, telemetry, and protocol-independent packet processing [8,9,17,18]. These capabilities have been explored in IIoT and industrial-network contexts [19,20]. However, most programmable-network approaches remain network-centric: policies are expressed in terms of flows, addresses, ports, or slices, rather than through standardized industrial asset semantics.

Industrial cybersecurity research has also investigated digital-twin-based anomaly detection, situational awareness, and cross-layer monitoring. Examples include digital-twin intrusion detection in industrial control systems [21,22], wide-area situational awareness for critical infrastructures [23], and bi-anomaly detection for Industry 4.0 environments [24]. Machine-learning approaches, including reconstruction-based anomaly detection [25] and malware classification [26], can improve detection coverage. Nevertheless, these approaches often remain monitoring-oriented: they detect or classify abnormal behavior, but they do not necessarily close the loop by translating asset-level observations into programmable network actions.

5G connectivity further increases the relevance of asset-aware network control. Industrial 5G supports remote maintenance, quality-of-service management, and network slicing [27,28]. Related work on semantic slicing, edge orchestration, and multi-layer resource management shows how application requirements can guide network and compute-resource allocation [29–33]. These directions are aligned with application-aware and deterministic networking efforts such as IETF APN and DetNet [34–36]. However, the

https://doi.org/10.3390/fi18070347

intent is typically represented through network-native or application-native abstractions, not through AAS-based industrial asset models.

Industry 5.0 broadens the design criteria for industrial digitalization. Networked industrial systems should not only support automation and connectivity, but also resilience, human-centric operation, and sustainability. Human digital twin and human-centered production studies show how digital representations can include operators, work contexts, and assisted decision-making [37,38]. Sustainability-oriented digital-twin work and ITU-T recommendations on network energy and carbon assessment provide a foundation for carbon-aware digital infrastructures [39–42]. These directions motivate network-control mechanisms in which asset state, operator intent, and operational constraints can influence connectivity decisions.

Table 1 summarizes the main areas of related work and the gap addressed by the proposed architecture.

Table 1. Related-work areas and identified gaps.

Related-Work Area	Current Limitation	Role of This Work
AAS interoperability	AASs standardize asset semantics, but rarely expose network-control operations.	Models network control and data planes as AAS-managed assets.
Programmable networking	SDN/P4 enables runtime enforcement, but remains mostly network-centric.	Translates AAS-level operations into SDN/P4 actions.
Industrial security monitoring	Digital-twin security often stops at monitoring or alerting.	Couples AAS telemetry with in-network counters for mitigation.
5G-enabled maintenance	Remote maintenance increases OT exposure and requires coordinated access control.	Coordinates 5G access, OT isolation, and rollback through AAS operations.
Industry 5.0 operation	Resilience, human-centricity, and sustainability require network decisions to consider operational continuity, operator control, and environmental constraints.	Shows how AAS-driven network control can support Industry 5.0 pillars.

The closest related approaches can be compared along five dimensions: whether they rely on standardized AAS semantics, whether they support SDN/P4-based enforcement, whether monitoring and enforcement form a closed loop, whether 5G connectivity is part of the architecture, and whether validation includes industrial hardware-in-the-loop. Table 2 reports this comparison.

Table 2. Positioning of this work with respect to closely related approaches. ● indicates full support; ◐ indicates partial support or conceptual alignment without full end-to-end implementation; ○ indicates absence of the capability.

Approach	AAS-Driven	SDN/P4 Enforcement	Closed Loop	5G Integration	HIL Validation
Digital-twin IDS for ICS [21]	○	○	◐	○	○
Digital-twin-based ICS security monitoring [22]	○	○	◐	○	○
AAS security descriptors [15]	●	○	○	○	○
AAS-based IAM and access control [43]	●	○	○	○	○
P4-based IIoT enforcement [17]	○	●	◐	○	○
Programmable networking for ICS [19]	○	●	◐	◐	○
Zero-trust architecture for IIoT [44]	○	◐	○	○	○
Semantic 5G slicing [29]	○	◐	◐	●	○
Edge resource marketplace [31]	○	◐	◐	●	○
This work	●	●	●	●	●

https://doi.org/10.3390/fi18070347

The resulting gap is the absence of an end-to-end mechanism that binds standardized industrial asset semantics to programmable network behavior. Existing AAS work provides asset representation, programmable-network work provides enforcement mechanisms, and industrial-security work provides monitoring and detection methods. The contribution of this paper lies in coupling these layers operationally: AAS operations become triggers for SDN/P4 enforcement, while in-network counters are exposed back as AAS properties and correlated with asset-level telemetry. This bidirectional coupling is the distinguishing element of the proposed architecture.

3. Problem Statement

Modern OT environments are increasingly connected to IT services, edge platforms, and remote maintenance infrastructures. This connectivity improves visibility and operational flexibility, but it also exposes industrial assets that often rely on legacy protocols, static segmentation, and limited endpoint protection. In this setting, network security cannot be based only on host-level mechanisms or manually configured forwarding rules. The central problem addressed in this paper is the lack of an operational interface able to translate industrial asset semantics into programmable network behavior.

Existing AAS-based approaches provide a standardized representation of industrial assets, including their properties, services, and lifecycle information. However, they do not generally expose the communication network itself as an AAS-managed asset, nor do they define reusable mechanisms through which asset state, maintenance intent, or telemetry can directly influence network enforcement. Conversely, SDN and programmable data planes provide runtime control over forwarding behavior, but their policies are usually expressed through network-level abstractions such as addresses, ports, tables, and rules. The research problem, therefore, lies at the boundary between these two layers: how to bind standardized asset semantics to programmable network enforcement in a way that remains compatible with industrial interoperability requirements.

The design objective of this work is to make the OT network both programmable and asset-aware. Programmability is provided by SDN/P4 mechanisms capable of installing, removing, and updating forwarding rules at runtime. Asset awareness is provided by AASs, which expose the operational context required to decide when and how the network should be adapted. The proposed artifact is an AAS-driven network control abstraction in which network control-plane and data-plane functions are represented through dedicated AASs, while AAS operations are translated by a Network Manager into concrete SDN/P4 actions.

The methodological framing is Design Science Research (DSR), since the contribution is an engineered artifact that is designed, implemented, and evaluated in a representative technical environment [10,11]. Following the DSR process proposed by Peffers et al. [11], the study is structured as follows:

1. Problem identification: the absence of a semantic interface between AAS-based asset representations and programmable network enforcement.
2. Objective definition: the design of a network-control abstraction able to coordinate asset state, communication requirements, and security policies through AAS operations.
3. Design and development: the implementation of network-related AASs, operation handlers, and SDN/P4 integration mechanisms using existing open-source components.
4. Demonstration: the execution of representative use cases covering runtime traffic control, 5G-assisted maintenance, and cross-layer anomaly mitigation.
5. Evaluation: the assessment of the artifact on a hybrid virtualized and hardware-in-the-loop testbed, including latency measurements and functional validation of the control loop.

https://doi.org/10.3390/fi18070347

6. Communication: the presentation of the artifact, its validation results, and its limitations in this manuscript.

DSR is appropriate for this study because the phenomenon under investigation is created by the artifact itself: AAS-driven network enforcement does not exist as an observed organizational process, but as a designed mechanism that must be instantiated and evaluated. Alternative methodological frames would be less suitable. Action research would require iterative intervention in a specific industrial organization, while a case-study methodology would primarily explain an existing real-world setting. In contrast, DSR directly supports the design–build–evaluate cycle required to validate the proposed AAS-to-network control abstraction.

The resulting artifact addresses the problem by introducing network-related AASs that operate alongside the AASs of OT assets. These AASs expose network-control operations and telemetry properties through standardized asset-oriented interfaces. The Network Manager interprets these operations, translates them into SDN/P4 policies, and installs the corresponding rules in the programmable data plane. In the opposite direction, network measurements such as counters and forwarding-state information are exposed back through the AAS layer and can be correlated with asset-side metrics.

The novelty is therefore not the isolated use of AAS, SDN, P4, or 5G, but their operational coupling. Asset state, maintenance intent, and cross-layer telemetry become inputs to network-control decisions, while network behavior becomes observable and controllable through the same semantic layer used to represent industrial assets. This coupling provides the basis for the threat model and validation use cases developed in the following sections.

3.1. System & Threat Model

We describe the system and threat model considered in this work through the layered ICS structure of the Purdue Enterprise Reference Architecture [45], as shown in Figure 1. This model organizes industrial networks into hierarchical levels, ranging from physical processes (Level 0) to enterprise systems (Levels 4 and 5). We focus on the OT layers, including sensors and actuators (Level 0), basic control systems such as Programmable Logic Controllers (Level 1), supervisory systems such as SCADA and HMIs (Level 2), and site operations (Level 3).

In our system model, we assume that the AAS instances used for monitoring are securely instantiated with integrity protections in the IT network (Level 4/5). To bridge the gap with the OT world, each AAS maintains a dedicated, secure channel to its specific physical asset, enabling monitoring while remaining logically separated from the industrial control loop. This assumption is intentionally explicit. The prototype evaluates whether trusted AAS and controller components can translate asset semantics into network enforcement; it does not claim that the AAS infrastructure is intrinsically non-compromisable. In a production deployment, the AAS servers, registry, SDN controllers, and policy adapters must themselves be protected with authentication, authorization, logging, network segmentation, and integrity monitoring consistent with zero-trust principles [44].

3.2. Adversary Model

With respect to the attacker model, we consider an internal or adjacent adversary who has obtained access to the OT network and possesses partial Dolev–Yao [46] capabilities. A typical Dolev–Yao attacker has full control over the communication medium, allowing them to access, modify, inject, delete, or delay any message passing through the network. While this adversary possesses almost unlimited network capabilities, he is inherently constrained by the perfect cryptography assumption. This implies that the attacker cannot break cryptographic primitives and can only decrypt a ciphertext or forge a signature if

https://doi.org/10.3390/fi18070347

he possesses the correct keys. Consequently, computational attacks such as brute-force private-key attacks or dictionary attacks cannot be performed.

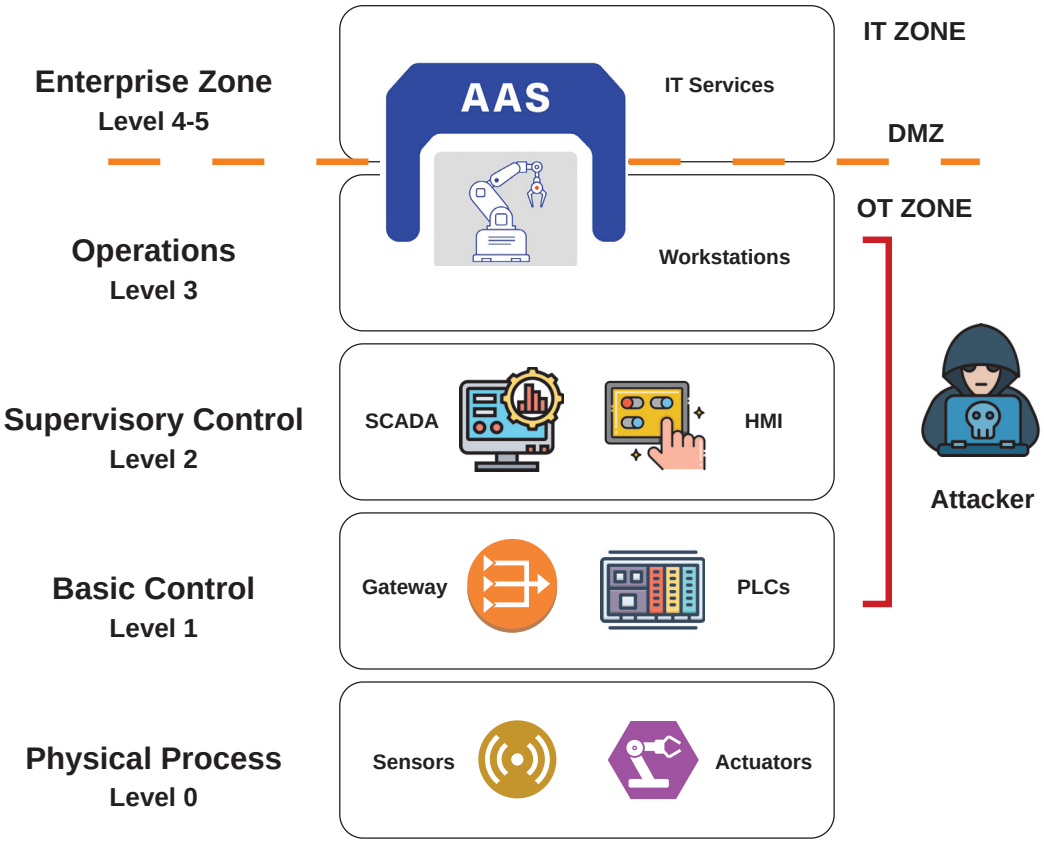


Figure 1. Threat model context based on the Purdue Reference Architecture. The attacker operates only within Levels 1–3. The Asset Administration Shell is trusted outside the attack surface.

Additionally, we introduce two specific constraints relevant to our ICS scenario:

- Aligning with previous works in ICS security [47–50], we assume that the attacker is able to operate only at or above Level 1 of the Purdue Enterprise Reference Architecture.
- We assume that the communication channel between the AAS and its physical asset is secured [15,43]. Since the adversary is inherently bound by the perfect cryptography assumption, he cannot compromise the encryption or integrity of this channel and therefore cannot tamper with the AAS monitoring data stream.

Finally, given the adversary's nature as a network-based Dolev–Yao intruder constrained by perfect cryptography and operating strictly within the OT network boundaries (Levels 1–3), he is inherently prevented from performing the following actions:

- The adversary cannot modify the AAS internal logic, APIs, or data schemas. Since the AAS resides in the protected IT domain and the Dolev–Yao model restricts the attacker to manipulating network traffic rather than executing arbitrary code on remote endpoints, the AAS's internal state remains inaccessible. So, the AAS is not part of the attack surface.
- Due to the perfect cryptography assumption, the adversary cannot decrypt traffic flowing through the secure tunnel between the AAS and the asset, nor can he inject valid messages into this channel without the necessary session keys.
- Since the adversary operates strictly above Level 1 and lacks physical access to the facility, he cannot manually tamper with hardware (e.g., cutting wires) or directly flash malicious firmware onto field devices.

https://doi.org/10.3390/fi18070347

3.3. Attack Objectives

The attacker's primary goal is to manipulate or degrade process integrity by exploiting weak authentication, insecure protocols, or poor segmentation between industrial levels. We identify three main attack classes:

1. Traffic Tampering: Includes MITM attacks on industrial protocols (e.g., Modbus, MQTT) to alter control logic, telemetry, or state updates between PLCs and sensors.
2. Sensor Data Spoofing: Involves falsifying telemetry readings (e.g., spoofed temperature, vibration) to mislead control systems or bypass safety thresholds.
3. Actuator Command Injection: Involves the unauthorized issuance of actuation commands (e.g., open valve, stop conveyor), leading to unsafe or unintended process changes.

Figure 2 shows an abstract attack tree modeling how an attacker at Level 1–3 might affect system integrity. The root node represents the goal (disruption), and leaf nodes describe specific tactics.

Finally, Table 3 summarizes the attacker's potential capabilities at each level, and whether the AAS can observe such activity given its architectural placement.

Table 3. Adversary capabilities by ICS level and AAS observability.

ICS Level	Attacker Capability	Observable by AAS
Level 3	Network traffic injection/replay	Yes (via network interface monitoring)
Level 2	SCADA/HMI traffic tampering	Partially (if logs or traffic mirrored to AAS)
Level 1	Sensor spoofing/ actuator injection	Indirectly (via inconsistent telemetry or command timing)
AAS/controller infrastructure	Trusted in this model	Security events only if assumption is relaxed

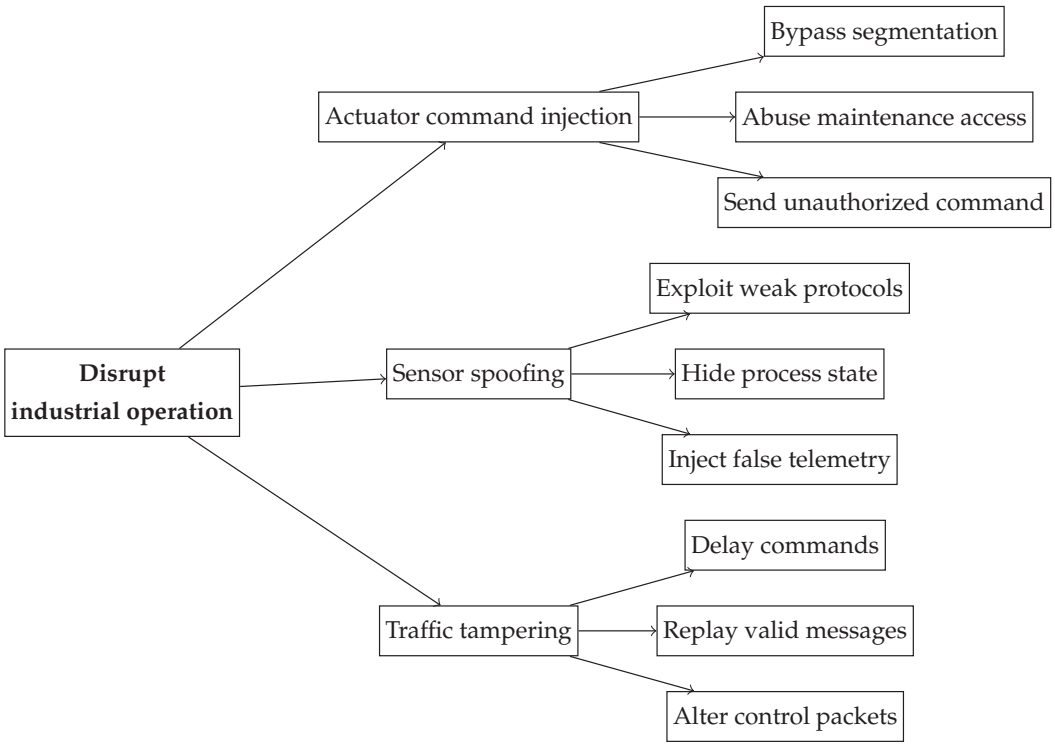


Figure 2. Attack tree for a Level 1–3 adversary attempting to disrupt industrial operation.

4. Architecture

This section presents the architecture that realizes the proposed asset-aware closed-loop network control abstraction for industrial environments. Rather than treating the communication infrastructure as a passive substrate, the architecture elevates the OT network to an industrial digital asset by equipping both its control and data planes with dedicated Asset Administration Shells. Through this design, asset semantics are mapped at runtime to concrete network enforcement actions implemented via programmable data planes. In parallel, in-network telemetry is continuously collected and exposed through the network AASs, where it is correlated with asset-level operational metrics. This bidirectional semantic coupling closes the control loop between assets and the network, as shown in Figure 3.

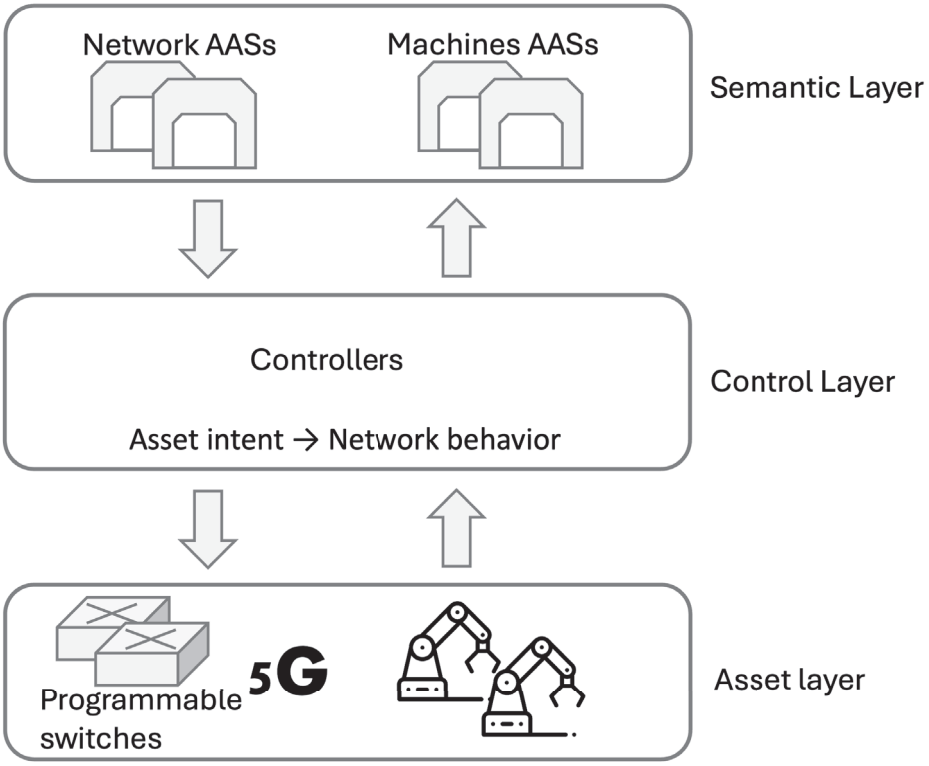


Figure 3. Illustration of the proposed control abstraction in which the OT network becomes an I4.0 asset through its own Asset Administration Shell (AAS). Asset-level operational state and intent are exposed via standardized AAS submodels and translated into network policies enforced at runtime through programmable data planes (SDN/P4).

4.1. Architectural Components

The architecture is shown in Figure 4, and its building blocks are:

- Three Factory Machines, i.e., the OT environment where actual production takes place.
- Two switches implementing the OT network data plane, thus allowing communication between the machines.
- A 5G network infrastructure, which we assume to be public and therefore usable to access the OT environment from the outside of the factory, i.e., for remote maintenance purposes.
- Two network controllers that can be used to program the network data plane, i.e., the switches' forwarding capabilities, as well as the interfaces towards the 5G network.
- The AAS infrastructure, which includes the various AASs needed, as explained later, and the AAS management infrastructure, typically hosted outside the OT in the factory IT network.

https://doi.org/10.3390/fi18070347

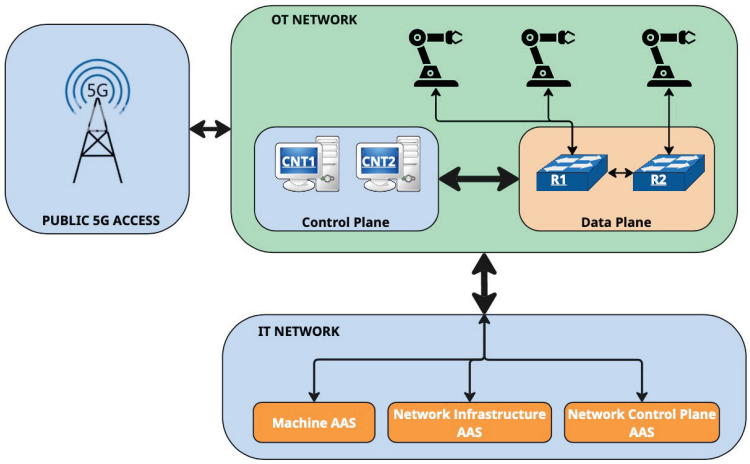


Figure 4. The proposed architecture that includes both the AAS Project and the testbed use cases.

In our testbed, the AAS infrastructure was implemented using Eclipse BaSyx [51], a set of open-source software components to support developers in implementing Industrie 4.0 concepts. Besides the AAS software tools, BaSyx provides the management infrastructure required to operate the AASs at runtime. This includes the AAS registry, a directory-like service through which AASs are announced and discovered, and the BaSyx web-based user interface. Each AAS is interconnected with the others and can be inspected or invoked by the administrator from the BaSyx web interface: the prototype interface views are reported in Appendix A (Figures A1 and A2).

As previously explained, the network has its own AAS: following modern architectural principles, we decided to separate the control plane from the data plane. This separation is known as Control–User Plane Separation (CUPS) [52], a principle that emerged in LTE [53] and 5G [54] and is nowadays a best practice in network design, because it improves scalability and flexibility. Therefore, we designed two distinct AASs to model the network:

1. The control plane AAS, which pairs with the network control plane and therefore with the logic governing network behavior;
2. The network infrastructure AAS, which pairs with the network data plane, i.e., the network nodes, and therefore determines their behavior according to the logic dictated by the control plane.

The control-plane AAS exposes controller-selection and forwarding-rule operations, while the data-plane AAS exposes switch-level functions such as counter retrieval. The 5G capability is integrated as a submodel of the machine AAS because connectivity is asset-related in the maintenance scenario.

Finally, we highlight that in practice, neither of these two AASs directly manipulates network devices. Instead, they act as a semantic and operational abstraction layer, exposing high-level intent (e.g., isolate an asset, enable maintenance mode, retrieve traffic counter) through standardized submodels and operations.

The AAS submodels exploit SDN controllers to convert AAS-level intent into network-level forwarding rules and telemetry queries. These latter actions can nowadays be implemented with two different technological approaches:

1. OpenFlow: we implemented an AAS submodel to interact with the Ryu SDN controller (https://ryu-sdn.org, accessed on 22 June 2026) REST APIs;
2. P4 Runtime and P4: we implemented a custom controller capable of installing into the switches specific P4 programs exploiting the P4 Runtime API (https://github.com/p4lang/p4runtime, accessed on 22 June 2026).

https://doi.org/10.3390/fi18070347

In our testbed, we tested both approaches and obtained results that were quite similar in terms of interaction efficiency and response time. Nonetheless, P4 aligns better with current network programmability trends and provides more effective telemetry features and forwarding control. Discussing the specific differences between these two SDN approaches is beyond the scope of this manuscript; an extensive survey on the topic can be found, for instance, in [55].

Finally, we also implemented the AAS for the 5G mobile terminal (the so-called User Equipment, or UE), with the capability to control 5G connections, as discussed in detail in Section 6.2.

4.2. AAS–SDN Implementation, Integration and Control Loop

Figure 5 illustrates the sequence of interactions triggered by an AAS operation such as `EnableMaintenanceMode`.

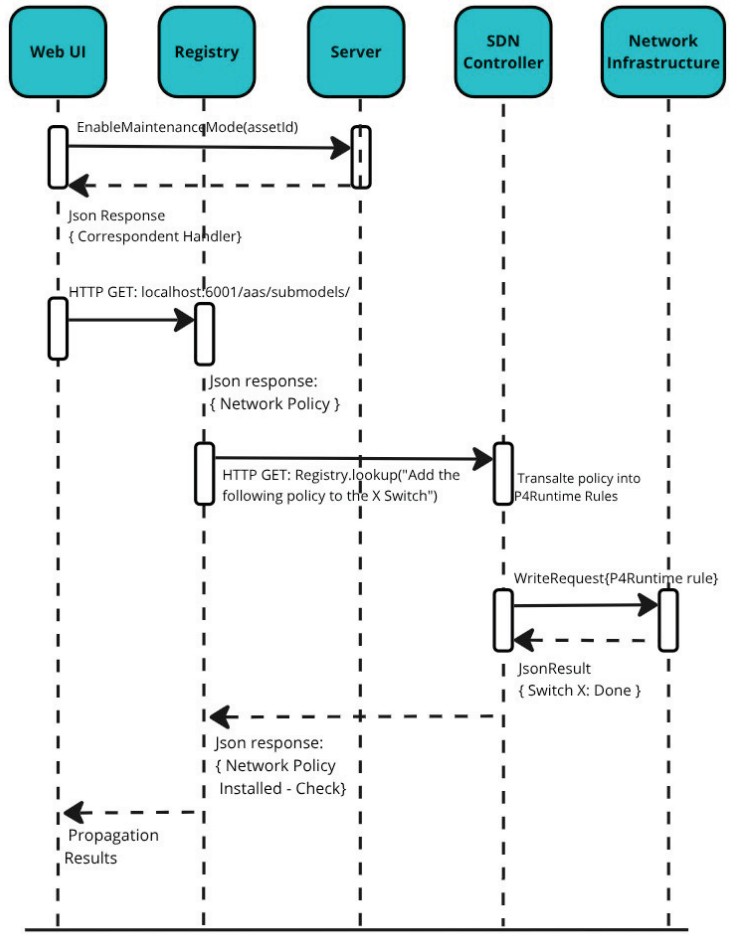


Figure 5. AAS-to-SDN operation workflow through P4Runtime.

The execution flow is as follows:

1. A network administrator invokes an AAS operation (e.g., `EnableMaintenanceMode(assetId)`) via the BaSyx Web UI.
2. The BaSyx Server dispatches the request to the corresponding AAS operation handler.
3. The handler invokes a policy adapter module, which formats the request as a network policy and forwards it to the SDN controller through a northbound API (REST or gRPC).
4. The SDN controller translates the policy into a set of P4Runtime `WriteRequests`, installing or updating table entries in the target P4 switch.

https://doi.org/10.3390/fi18070347

5. The switch enforces the new forwarding behavior immediately, while the AAS updates its exposed state properties (e.g., maintenance mode status and last update timestamp).

In the maintenance mode use case, this translation blocks intra-factory (east–west) traffic involving the selected asset, while allowing only explicitly authorized north–south traffic via the 5G interface. Security and connectivity policies are represented in the AAS through a dedicated Connectivity and Security submodel that encodes high-level intent independently of the underlying SDN technology. Each policy specifies scope and identifiers, traffic selectors (source/destination IP, protocol, port), the intended action (e.g., `ALLOW`, `DENY`, `MAINTENANCE_MODE`), the enforcement point, and lifecycle metadata. An excerpt of the submodel template used in the implementation is shown in Listing 1.

Listing 1. Example AAS connectivity security submodel (simplified).

```
{
  "idShort": "ConnectivitySecurity",
  "submodelElements": [
    {
      "idShort": "Policy_Block_M1_to_M2",
      "modelType": "SubmodelElementCollection",
      "value": [
        {"idShort": "policyId", "value": "blk-m1-m2"},
        {"idShort": "action", "value": "DENY"},
        {"idShort": "srcIp", "value": "10.0.1.10"},
        {"idShort": "dstIp", "value": "10.0.1.11"},
        {"idShort": "protocol", "value": "ANY"},
        {"idShort": "enforcementPoint", "value": "R1"},
        {"idShort": "status", "value": "INSTALLED"}
      ]
    }
  ]
}
```

In the prototype, policy updates are serialized by the Network Manager operation handler before they reach the SDN controller. This avoids simultaneous writes from multiple AAS instances in the evaluated use cases.

At runtime, the SDN controller maps this representation to concrete data-plane rules, e.g., by installing P4 table entries that match the specified fields and execute either forwarding or drop actions.

To support anomaly detection and situational awareness, the architecture integrates in-network telemetry with AAS-based observability. Traffic counters are maintained directly within the P4 data plane, either per-port or per-flow. The `GetCounters` AAS operation retrieves this data indirectly through the SDN controller:

1. The AAS invokes `GetCounters()` on the Network Infrastructure AAS.
2. The operation handler requests telemetry data from the SDN controller.
3. The controller issues a P4Runtime `ReadRequest` to retrieve the relevant counter values from the switch.
4. Retrieved statistics (packet and byte counts with timestamps) are returned to the AAS and exposed as read-only properties.

These properties are visualized in the Web UI and correlated with asset-level metrics (e.g., CPU usage from machinery AASs), enabling the distinction between transient traffic fluctuations and sustained anomalous behavior, as discussed in Section 6.

The main graphical interface for interaction is accessible through the IT Network and provides network administrators with all AASs, each accompanied by its respective submodel and operations. Each of them comprises multiple submodels needed to support the functions of the machinery or network component, focusing on network forwarding

https://doi.org/10.3390/fi18070347

rules and properties to achieve reliable, efficient, and secure communications within and outside the OT Network.

5. Testbed Setup

The testbed used to analyze the feasibility, effectiveness, and performance of the proposed artifact has a hybrid structure. The baseline environment is containerized and virtualized, while the hardware-in-the-loop extension replaces selected emulated elements with industrial-grade Revolution Pi equipment. It comprises three main components: the AAS Infrastructure and coordination system, the network configuration, and the 5G Infrastructure for external connectivity.

As already said, the AAS Infrastructure has been developed using Eclipse BaSyx. We used the BaSyx Registry (v1.5.0), BaSyx Server (v1.5.0), and AAS Web UI (v230703) to construct the main AAS distributed infrastructure, and the BaSyx Java Library to implement and deploy each AAS.

The AASs are hosted on the IT network using an Apache HTTP Server (v2.4.57) and registered within the BaSyx Registry and BaSyx Server. Subsequently, the Web UI retrieves information about submodels, names, and addresses of all AAS instances registered in the Server and Registry components, enabling direct interaction via the user interface. This approach abstracts from low-level complexity irrelevant to the user, who, in the given use case, is only concerned with monitoring the security of Factory Machinery.

- Each Asset Administration Shell consists of one or more submodels.
- Each submodel provides one or more Operations or Properties: Operations allow the user to modify system parameters or specific factory components, whereas Properties display relevant data about a given component.
- Each Asset Administration Shell can connect with others based on the provided network configuration, thereby contributing to overall factory coordination and security.

At the other end, the primary tool used to deploy all components and configure the Network in a single virtualized environment is Kathará [56] (v3.8.3), an open-source container-based network emulation system designed for demonstrating, developing, and testing production networks in a controlled sandbox. Kathará employs Collision Domains as fundamental network nodes. More specifically:

- Each component requiring network access is linked through one or more collision domains;
- Every component in this setup is connected to specific collision domains, ensuring that each P4 Switch (R1 and R2) can exchange information with Network Controllers, the AAS Infrastructure, and individual Machinery.
- Furthermore, each component connected to a collision domain must be assigned a unique IP address and a distinct network route configuration;
- Finally, each machinery unit is equipped with two different IP interfaces to its corresponding SDN switch. This setup simulates a realistic factory environment where machinery maintains multiple connections to the factory network and other components.

5.1. Hardware-in-the-Loop

To move beyond a purely software simulation, we extended the architecture with Hardware-In-The-Loop (HIL). As shown in Figure 6, the reference topology contains three central machines attached to the factory network and two P4 switches with external controllers, CNT1 and CNT2. In the HIL configuration, one machine and one SDN switch are replaced by Revolution Pi equipment (https://revolutionpi.com/en, accessed on 22 June 2026), as shown in Figure 7. Revolution Pi combines a Linux-based software environment with industrial-grade deployment characteristics. The overall architecture remains unchanged, as does the network configuration, thanks to Kathará's ability to connect to

https://doi.org/10.3390/fi18070347

external devices through Linux Bridges. The main RevPi is still connected to its corresponding collision domain as before. Still, it can now provide and gather information from sensors and actuators attached to it in a more standardized, industrial-friendly manner, thereby enabling the operability of Factory Machines.

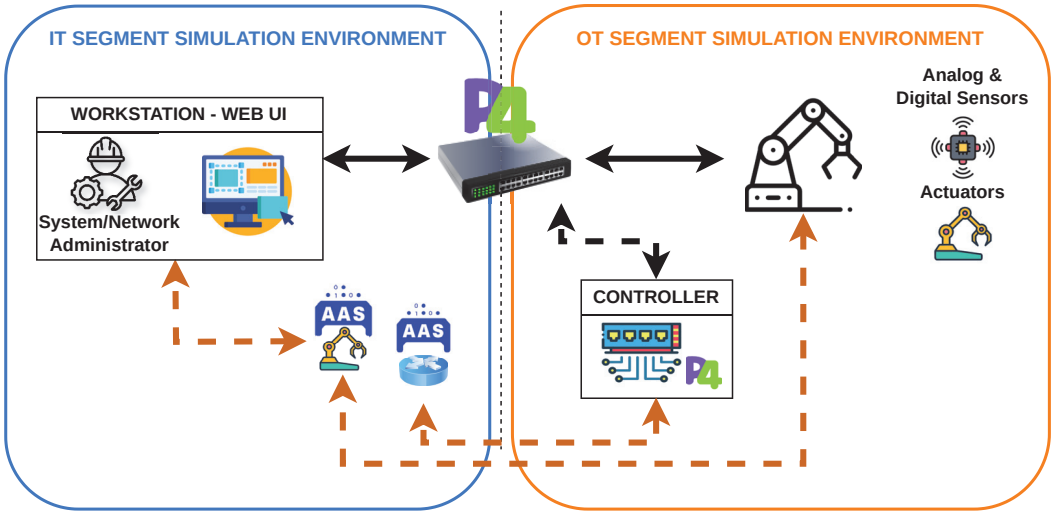


Figure 6. Containerized testbed infrastructure and collision domains.

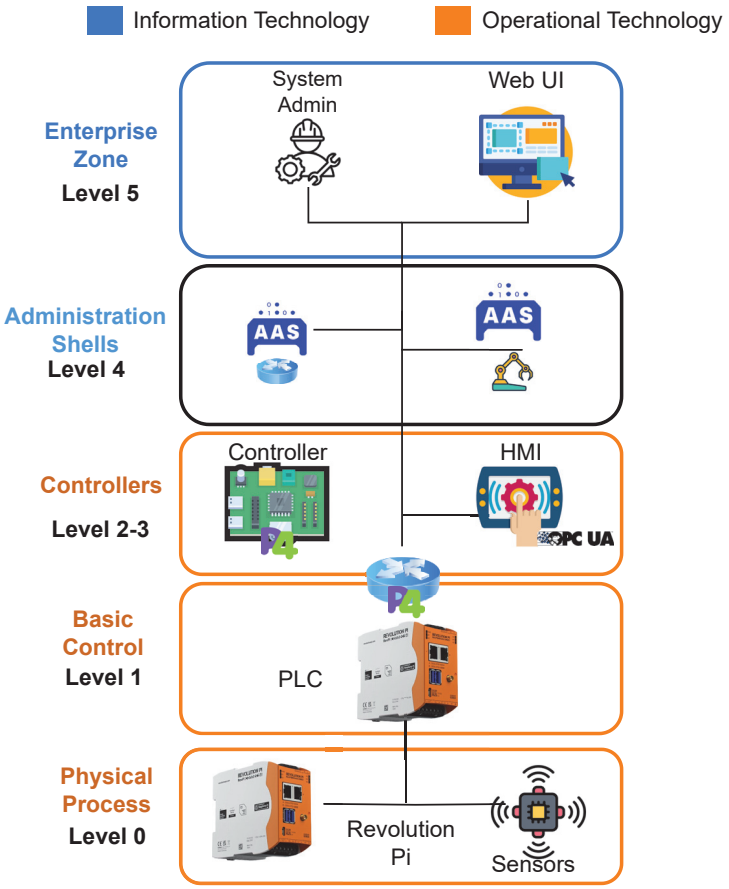


Figure 7. Hardware-in-the-Loop extension with Revolution Pi equipment.

5.2. 5G Implementation

The last step was to test the proposed architecture with external connectivity capabilities. Because access to an industrial private 5G deployment was outside the reproducible software testbed, we implemented a virtualized 5G infrastructure by means of:

- *UERANSIM* (https://github.com/aligungr/UERANSIM, accessed on 22 June 2026) which provides an open source 5G UE and RAN (gNodeB) implementation;
- *Open5GS* (https://open5gs.org/, accessed on 22 June 2026) which provides an open-source 5G core network implementation, including all the various services and supports a complete separation between the control and data plane.

Then we implemented an AAS to control 5G connectivity, including the capability to open 5G PDU sessions to specific destinations and to control QoS (PDU session minimum available bit rate). With these components, it is possible to enable 5G Communication from and to all network hosts in the Kathará virtualized structure that have a virtualized UE (User Equipment). This allows for different communication media, representing a more varied and standardized way to control, perform maintenance, or remotely access factory infrastructure.

6. Results

This section presents a set of use cases designed to validate the proposed architecture. The goal is to demonstrate the capability of the AAS to automate network control and enforce dynamic security policies. Specifically, we focus on the system's ability to customize forwarding rules based on operational needs, manage secure remote maintenance directly through the standard AAS interface, and isolate compromised assets.

The experimental validation relies on the reference architecture depicted in Figure 8, which comprises three manufacturing machines, two programmable switches implementing the P4 Runtime specification [57], and two custom network controllers developed in Go (https://www.go.dev). The infrastructure is logically partitioned into a control plane, hosting the decision-making logic, and a data plane, consisting of the forwarding devices.

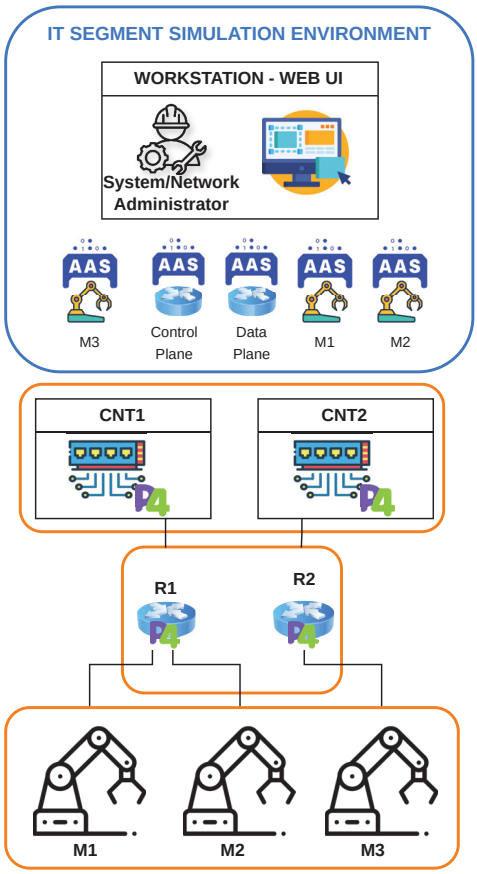


Figure 8. Reference topology employed in the validation use cases.

6.1. Real-Time Customization of Traffic Forwarding Rules

The first outcome of this workflow is illustrated in Figure 9. It shows how traffic within the OT network can be managed and customized in real time through interactions with the Control Plane and Data Plane AASs. Initially, a continuous, high-speed traffic stream is generated, with 1512-byte packets. Subsequently, the network manager activates and configures the SDN controller managing the P4 Switch to enforce a policy that blocks communication between the involved machines. This action is initiated via the AAS and is represented in the figure by the first spike, corresponding to the initial exchange between the controller and the switch. As expected, traffic is successfully interrupted. Afterward, the network manager engages directly with the Network Infrastructure AASs to manually apply new rules: first to permit traffic, and then to block it again (the second and third spike, respectively). These two interactions reflect the communication between the BaSyx Web User Interface and the Control Plane's AAS inside the factory network infrastructure. At the final step, the controller is deactivated, restoring standard network functionality.

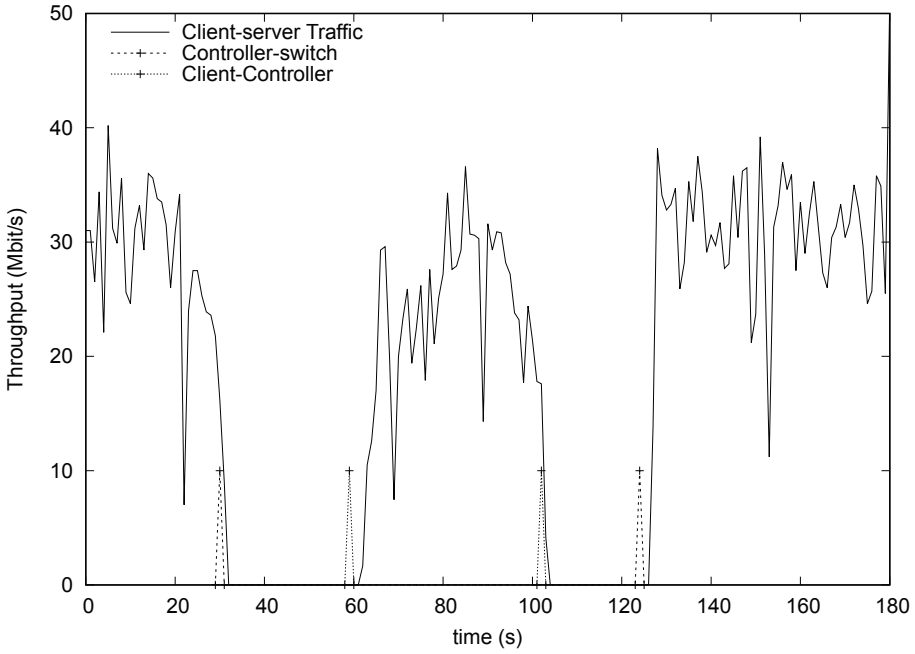


Figure 9. Example of the OT network traffic control achievable through the AAS. A packet flow between the involved hosts is blocked. The policy is then modified by interacting with the AAS submodels and restoring the traffic flow.

To provide a quantitative assessment, the experiment described in this first use case on real-time traffic control, depicted in Figure 9, was repeated 1000 times. Specifically, we evaluated the latencies associated with the rule-management operations (installation and removal) required to block and restore traffic flow. This large sample size provides a stable empirical estimate and allows for a precise decomposition of the total execution time. The results are summarized in Table 4.

Table 4. Statistical analysis of Control Plane latency metrics over 1000 iterations. Values are reported in milliseconds (ms).

Latency Component	Mean	Std Dev	Min	Max
End-to-End Request (T_{total})	4.2106	1.2731	2.7680	37.6330
Controller–Switch Interaction (T_{net})	1.9147	0.5260	1.2650	10.5500
AAS Overhead (T_{aas})	0.0397	0.0216	0.0170	0.5190

https://doi.org/10.3390/fi18070347

The End-to-End Request time, denoted as T_{total}, represents the aggregate latency perceived by the IT client (i.e., the AAS Web GUI), encompassing the entire command chain from issuance to acknowledgement. The Controller–Switch Interaction time (T_{net}) measures the latency required for the SDN controller to communicate with the P4 switch via P4Runtime to apply or remove the forwarding rules. This component (T_{net}) constitutes the baseline physical latency of the network infrastructure. Finally, the AAS Overhead (T_{aas}) isolates the computational cost introduced by the AAS logic, including request parsing and semantic adaptation.

The experimental data indicate that AAS integration introduces negligible overhead ($T_{aas} \approx 0.04$ ms), accounting for less than 1% of the total round-trip time. The system performance is dominated by the network baseline (T_{net}), which averages 1.91 ms. Regarding the operational impact, T_{net} represents the effective transition window during which the network configuration is updated. To place the observed values in context, the Industry 4.0 Digital Twin literature reports that real-time monitoring, sensor-driven synchronization, and control-loop stability require predictable round-trip times typically in the millisecond-to-tens-of-milliseconds range, depending on the closed loop being supervised [58–60]. The measured transition interval of approximately 2 ms is therefore compatible with supervisory security loops and with monitoring-grade synchronization between an asset and its AAS.

In the first use case, the primary objective is to demonstrate the feasibility of real-time monitoring of Network OT traffic and, as a direct consequence of network forwarding rules, subsequently modifying network behavior from a single interface within the IT Network. This highlights the connection between OT and IT Networks, a fundamental concept in Industry 4.0 that increases productivity, cybersecurity, and the overall efficiency of the entire plant. The primary user of this system, represented in Figure 8 as the Network Administrator, has the capability to modify the network setup to enable or disrupt communication between network nodes.

6.2. SDN-Based Maintenance Mode in a 5G-Connected Industrial Environment

This second use case models remote maintenance as a controlled human-in-the-loop operation. We assume the machines can be connected to an external 5G network, in order to be accessed remotely with a pre-defined quality of service. In this case, the interaction with the network infrastructure will be available only through the network terminal. We therefore implemented a 5G connectivity submodel in the AAS of each machine enabled for external access. The submodel controls the User Equipment and exposes operations to establish external connectivity through the public 5G network, in order to enable remote maintenance of factory machines.

Since external connections expose OT systems to attacks from the outside world, we consider it of paramount importance to control them in an integrated manner with internal connectivity to minimize the risk of cyberattacks. When a given asset is stopped and requires remote maintenance, we assume that the asset must first be disconnected from the OT network, in order to avoid potential backdoors into the production environment. It is then connected to the remote operator through a dedicated 5G PDU session.

The AAS 5G submodel interacts with the 5G User Equipment, i.e., the UERANSIM simulator in this case, and may request PDU session setup and tear-down. A screenshot of the related AAS and its Mobile Communication submodel is provided in Appendix A (Figure A3).

At the same time the P4 switch to which the asset is connected must block all traffic from/to the asset and the OT network (as shown in Figure 9), and the asset must then establish a PDU session towards the 5G. The transition is shown in Figure 10. The machine

https://doi.org/10.3390/fi18070347

first exchanges traffic on the internal OT network (purple line). The local OT traffic is then blocked, and the 5G traffic (green line) is started. After a configurable maintenance window, the original OT operations are restored. In the same figure, the purple plot is registered on the internal P4 switch and the green plot on the asset 5G interface, superimposed on the same time scale.

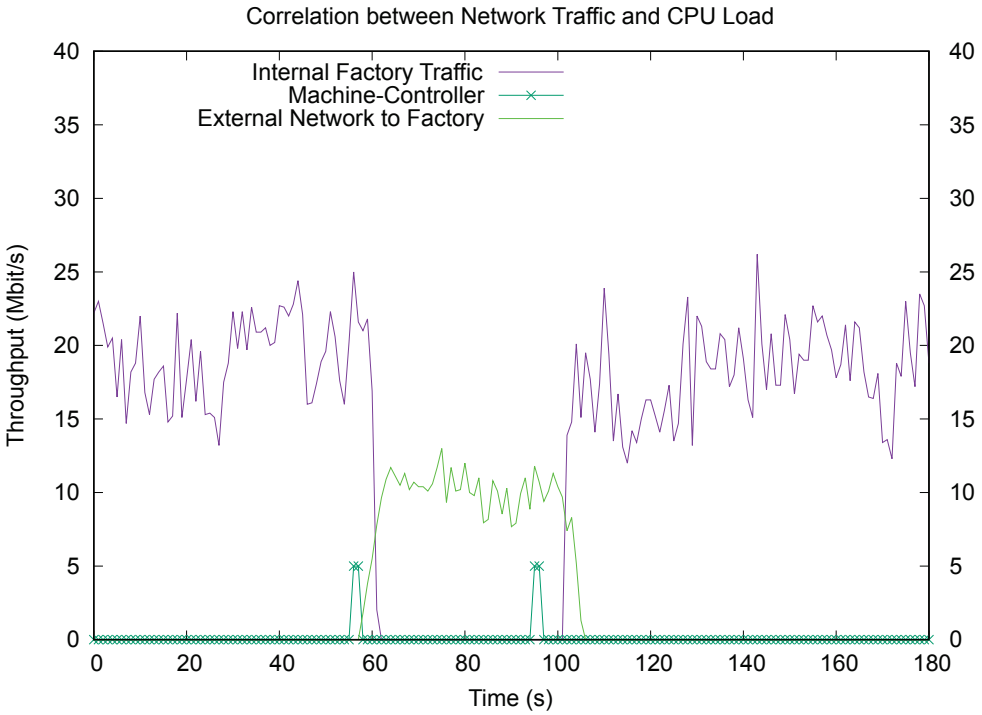


Figure 10. Maintenance mode transition between internal OT traffic and 5G access.

From an Industry 5.0 perspective, the maintenance mode is not merely an automated machine event. It is a safe interface for a remote expert or an on-site operator to interact with an asset while the AAS keeps the network state explicit, auditable, and reversible. The AAS therefore reduces cognitive and cybersecurity burden: the operator requests a semantic maintenance action, and the infrastructure coordinates isolation, external reachability, and rollback.

To provide some insight into this latter operation, we measured the time required to open and close the PDU session, as shown in Figure 11. These values are functional measurements from the virtualized 5G environment. They validate the orchestration chain, but should not be generalized to production radio latency without dedicated 5G hardware testing.

6.3. Anomaly Detection and Isolation via Cross-Layer Correlation

This third use case focuses on demonstrating cross-layer correlation in the context of remotely accessible OT networks, where exposure to external threat vectors increases the attack surface. Accordingly, our system implements a security mechanism based on cross-layer correlation, avoiding reliance on isolated network or machinery data. This combined analysis allows the system to distinguish between legitimate operational bursts and malicious activities.

First, the system monitors real-time network telemetry via packet–byte counters embedded in the P4 Switches. Specifically, the system utilizes the `GetCounters` operation, as detailed in Section 4, to periodically retrieve traffic statistics from the Network Infrastructure AAS. Then, this network data is not analyzed in isolation but is correlated with host-level metrics retrieved from the Asset Administration Shell of the target machinery.

https://doi.org/10.3390/fi18070347

In this specific use case, the chosen target machinery is the RevPi, integrated into the Hardware-in-the-Loop setup, as described in Section 5, while the system metric is the CPU usage, retrieved directly from the RevPi's AAS.

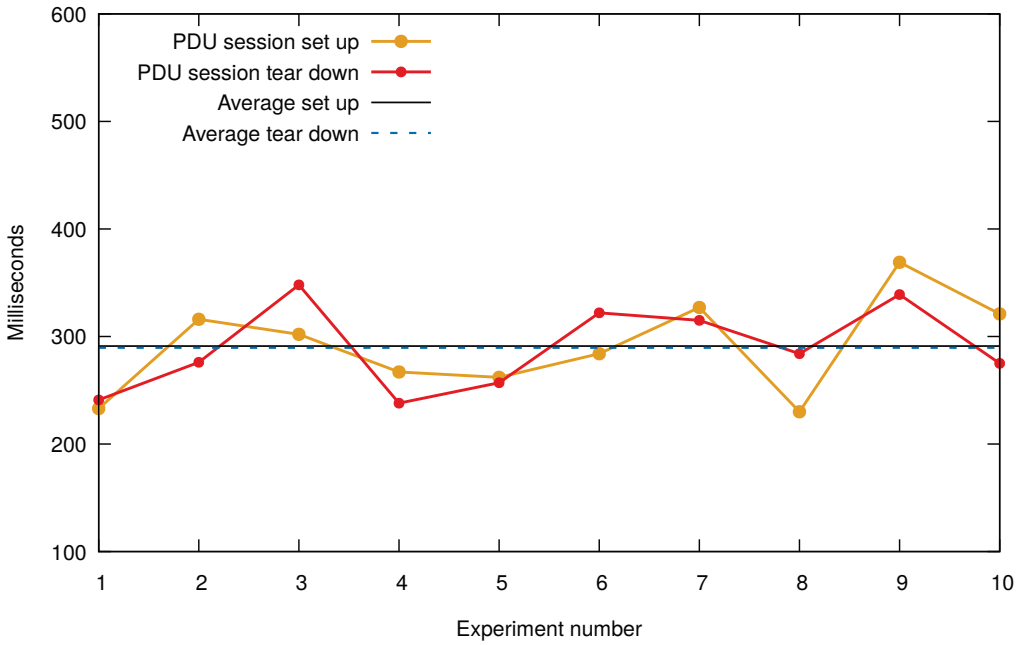


Figure 11. Time to set up and tear down the 5G PDU session. The plot shows 10 experiments and their respective averages.

6.3.1. Baseline Characterization and Threshold Definition

To ensure rigorous detection logic and minimize false positives, we conducted a 4 h preliminary profiling phase. During this period, we monitored the system under standard operational conditions to establish operational baselines and define specific safety thresholds for both network traffic and machinery performance. Telemetry data was retrieved from both AAS instances every second, and the observed metrics are summarized in Table 5.

Table 5. Operational baseline metrics for the target machinery (RevPi).

Metric	Mean Value	Max Recorded Value
CPU Usage	29.35%	51.68%
Network Throughput	3.89 Mbps	4.90 Mbps

Based on these observations, the following detection parameters were defined:

- RevPi Performance (CPU): while the RevPi exhibits an average CPU utilization of ≈29%, transient peaks can reach ≈51% during standard operations. To accommodate this normal load variance while identifying potential anomalies, the safety threshold was set at $\tau_{cpu} = 40\%$.
- Network Throughput: with a baseline traffic volume averaging 3.89 Mbps and legitimate spikes up to 4.90 Mbps, the network throughput threshold was conservatively set at $\tau_{net} = 10$ Mbps.

Crucially, the detection mechanism incorporates a temporal consistency check to mitigate false positives caused by transient spikes, such as legitimate load fluctuations observed in the RevPi's baseline. Given the sampling interval of one second, a momentary threshold

https://doi.org/10.3390/fi18070347

violation is insufficient to trigger an alarm. Instead, the system enforces a persistence constraint: an anomaly is formally declared only if the safety thresholds for both CPU usage and network throughput are simultaneously and continuously exceeded for a detection window of $W_{detect} = 30$ s. This logic filters operational noise in the evaluated scenarios. The parameters (τ and W_{detect}) were empirically calibrated for this testbed and should not be read as universal thresholds. More advanced models, including reconstruction-based anomaly detection and unsupervised learning, can be integrated at the AAS layer without changing the SDN enforcement path [25,61].

6.3.2. Attack Detection Scenarios

The detection workflow relies on the simultaneous violation of both τ_{cpu} and τ_{net}. The Network Manager submodel periodically polls the `GetCounters` property from the Network Infrastructure AAS and correlates it with the `CPU Usage` property from the Machine AAS. We illustrate two distinct scenarios to validate the robustness of this correlation mechanism against operational fluctuations that may be detected as false positives and its efficacy in mitigating actual threats.

First, we address the identification of a transient anomaly. Figure 12 illustrates a scenario where a temporary surge in traffic is detected, triggered by a legitimate system maintenance task: a software update initiated on the RevPi by the system administrator. This operation causes a simultaneous increase in network throughput (due to package downloading) and CPU usage (due to installation overhead). However, the violation of the thresholds is transient. Since the data values return to the nominal baseline ($<\tau_{net}$ and $<\tau_{cpu}$) within the predetermined observation window, the system correctly classifies this event as a non-malicious operational spike. Consequently, no blocking action is triggered, demonstrating the system's ability to maintain operational continuity and avoid unnecessary disruptions typical of overly aggressive intrusion detection systems.

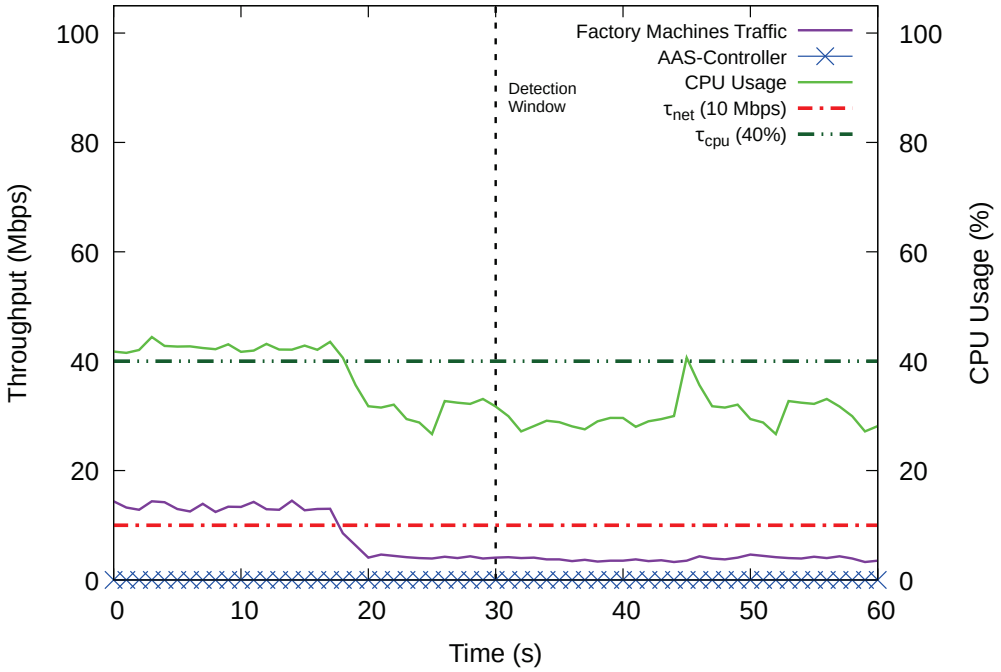


Figure 12. Transient anomaly scenario: the system detects a simultaneous spike in network throughput and CPU usage caused by a generic software update on the target machine. Since the values rapidly return to nominal levels, the correlation logic identifies this as a transient anomaly, and no isolation measures are taken.

https://doi.org/10.3390/fi18070347

Conversely, the second scenario demonstrates the detection and isolation of a sustained attack. Figure 13 depicts the system dynamics during a confirmed volumetric Denial-of-Service attack targeting the Modbus server on the RevPi, executed by an adversary operating from a compromised workstation located at Level 3 of the Purdue Enterprise Reference Architecture, as described in Section 3. In this context, the protection mechanism operates through a deterministic closed-loop sequence:

1. Monitoring and Correlation: the Network Manager observes a synchronized violation of both thresholds (τ_{net} and τ_{cpu}), indicating that the surge in network traffic is directly degrading the asset's computational performance.
2. Temporal Confirmation: unlike the transient update scenario, the violation persists continuously throughout the entire duration of the detection window (W_{detect}), satisfying the condition for anomaly confirmation.
3. Active Mitigation: upon confirming the threat, the Network Manager triggers an automated response via the SDN Controller AAS to enforce traffic isolation at the data plane level. Specifically, the controller sends a P4Runtime `WriteRequest` to update the P4 switch's IPv4 forwarding table (`ipv4_forward`). A new flow entry is installed that matches the malicious source IP address and executes a `drop` action. As evidenced in the graph, the throughput drops to operational values immediately after the detection window (approx. $t = 30$ s), confirming the successful isolation of the compromised host.

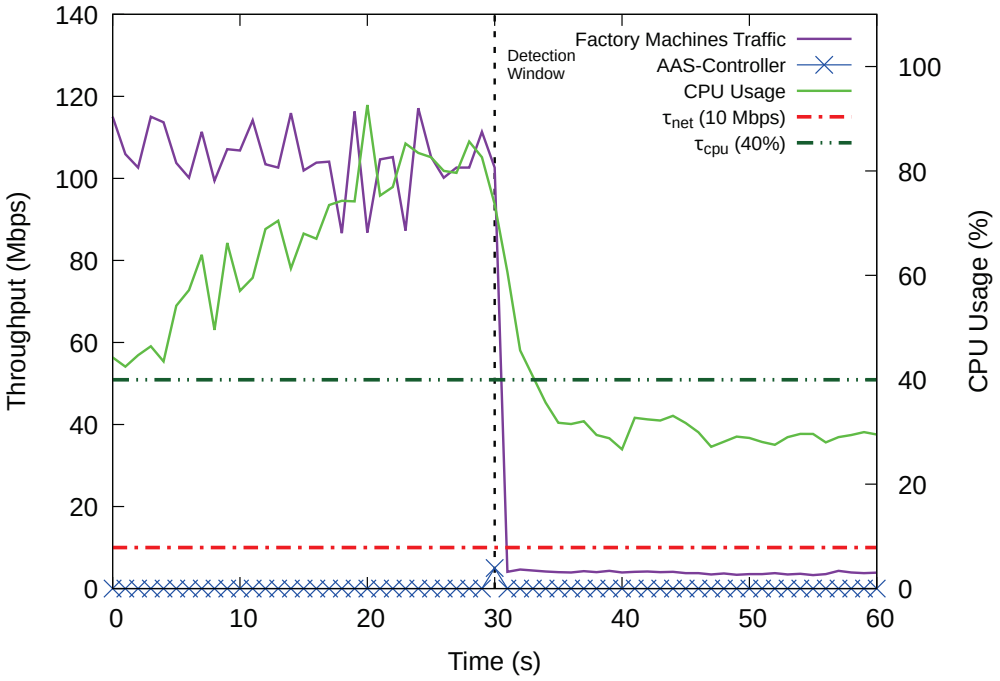


Figure 13. Anomaly scenario: a sustained violation caused by a volumetric DoS attack targeting the RevPi's Modbus server. The system correlates the high network load with abnormal CPU usage and triggers an automated response. The SDN controller pushes a blocking rule, effectively isolating the attacker.

Given the limited number of controlled scenarios, the results are interpreted as proof-of-concept evidence rather than as statistically grounded attack-success, false-positive, or false-negative rates. In the evaluated cases, the defense mitigated the sustained DoS scenario, while no isolation was triggered during the transient benign scenario. A broader quantitative assessment would require a labeled dataset covering multiple attack classes, benign operational and maintenance activities, and low-volume manipulations.

https://doi.org/10.3390/fi18070347

The implemented countermeasure relies on a binary drop action to validate the AAS-to-P4 enforcement loop with a minimal and deterministic mitigation policy. For Industry 5.0 resilience, the same mechanism can evolve toward graceful degradation: P4Runtime could install rate-limiting rules instead of drops, redirect suspicious traffic to an inspection path, preserve worker-safety telemetry, and maintain emergency human-override channels while isolating the malicious vector. The AAS layer is suitable for this extension because it can expose not only network addresses but also the asset role, maintenance state, and safety criticality needed for selective mitigation.

7. Discussion, Limitations, and Future Work

The experimental results validate the proposed AAS-driven control loop as a mechanism for translating asset-level information into programmable network actions. This section discusses the experimental boundaries of this result and examines how the same architectural principles can support more scalable, resilient, and human-centered industrial network control.

7.1. Technology Readiness and Evidence Boundaries

The implemented prototype is classified as TRL 4, corresponding to component and system validation in a laboratory environment. The hardware-in-the-loop integration with Revolution Pi equipment provides partial evidence toward TRL 5, since the AAS and SDN/P4 control loop interact with industrial-grade hardware in a relevant testbed. The 5G subsystem, however, is evaluated through UERANSIM and Open5GS rather than commercial 5G infrastructure, while the anomaly-detection logic is assessed through controlled scenarios. The current evidence, therefore, supports laboratory-scale architectural validation rather than field deployment readiness.

The measured 1000-iteration rule-management experiment shows that the AAS processing component is small compared with controller–switch interaction time in the implemented setup. This supports the claim that the semantic AAS layer is not the dominant latency source. It does not, by itself, prove negligible packet loss, jitter, or behavior under concurrent production traffic; those metrics require dedicated load experiments and are future validation targets.

To make the maturity assessment auditable rather than monolithic, Table 6 maps each architectural block to its current TRL and to the evidence that supports the assigned level. The composite system-level TRL is therefore reported as TRL 4 with partial movement toward TRL 5: the AAS layer, the SDN/P4 control loop, and the HIL asset are validated in a relevant testbed, while the 5G subsystem remains at TRL 3–4 because it is exercised only through virtualized RAN/Core components.

7.2. Detection Scope, Feedback Delay, and Control Stability

The anomaly-detection use case validates the cross-layer detection and enforcement architecture through a deterministic threshold-based logic. The detector correlates network-side indicators (traffic volume) with machine-side indicators (CPU load) and triggers mitigation only when the abnormal condition persists over a confirmation window. This design is suitable for sustained volumetric or resource-exhaustion events, where the cyber and physical symptoms are jointly observable. It is less suitable for stealthy manipulations, such as isolated malicious Modbus commands, low-rate process deviations, or false-data injection events that do not produce persistent changes in CPU load or throughput. The architecture does not depend on this specific threshold logic. The detector is placed above the SDN/P4 enforcement interface: it consumes AAS-exposed telemetry and, when an anomaly is detected, invokes the same mitigation procedure used in the prototype.

https://doi.org/10.3390/fi18070347

Therefore, the detection module can be replaced or extended without changing the AAS-to-SDN/P4 enforcement path.

Table 6. TRL assessment summary.

Block	TRL	Validation Evidence	Experimental Boundary
AAS semantic layer (BaSyx)	4–5	Rule-management benchmark over 1000 iterations; integration with controllers and HIL asset.	Evaluated in laboratory conditions, not in a multi-site industrial deployment.
SDN/P4 control loop	4–5	End-to-end policy installation through P4Runtime on emulated and HIL communication paths.	Tested on a limited topology with two switches and controlled traffic patterns.
HIL industrial asset (RevPi)	5	Industrial-grade controller running representative workloads and exposing AAS-mediated telemetry.	Validated on a single HIL asset rather than a heterogeneous production cell.
5G connectivity subsystem	3–4	Orchestration validated with UERANSIM and Open5GS.	RAN and Core components are virtualized; commercial 5G hardware was not used.
Anomaly-detection logic	3–4	Threshold-based detector evaluated on one sustained DoS scenario and one transient benign scenario.	No large labeled dataset or multi-class attack evaluation is included.
Composite system	4 ($\rightarrow$5)	Integrated AAS, SDN/P4, 5G-emulated, and HIL validation in a relevant laboratory testbed.	System-level readiness is limited by the virtualized 5G subsystem and controlled detection scenarios.

More advanced detection logic can be integrated at the same architectural point, but the choice of detector is deployment-dependent. Data-driven alternatives require stronger assumptions: reconstruction models and one-class methods rely mainly on representative nominal data [25,61,62], while supervised classifiers require labeled benign and attack traces, which are often unavailable in OT environments. Lightweight unsupervised methods, such as Isolation Forests, can reduce computational cost and improve interpretability, but their effectiveness depends on the separability of anomalies from normal operational transients [63]. Accordingly, the detector is not considered a generic replaceable block with equivalent alternatives. It is a deployment-specific component whose design depends on telemetry availability, computational constraints, and the class of attacks to be detected. The architectural interface, however, remains unchanged: the detector consumes AAS-exposed telemetry and triggers mitigation through the existing AAS-to-SDN/P4 enforcement path.

The resulting loop should be interpreted as a supervisory security loop, not as a real-time process-control loop. Its purpose is to detect abnormal cyber–physical conditions and enforce network-level mitigation, rather than to stabilize physical dynamics at sub-second time scales. The loop latency includes telemetry sampling, the confirmation window used by the anomaly detector, AAS request processing, controller-side policy handling, and P4Runtime rule installation. This timing is compatible with the mitigation of sustained security events, such as volumetric or resource-exhaustion attacks, but it is not appropriate for immediate actuation or stability-critical control. Future extensions should therefore evaluate shorter sampling intervals, adaptive confirmation windows, and mitigation policies explicitly designed for cyber–physical workflows requiring tighter synchronization.

7.3. Scalability

Scalability in the proposed architecture mainly concerns the AAS-based management path and the SDN/P4 enforcement path. An AAS operation is processed by the BaSyx infrastructure, dispatched to the corresponding operation handler, translated by the Network

Manager into a network policy, and finally installed by the SDN/P4 controller through P4Runtime. The OT data plane is involved only after the rule has been installed, through match–action processing in the programmable switch. Therefore, the AAS layer does not add per-packet processing overhead to industrial traffic; its impact is associated with the rate and concurrency of management operations. As the number of assets increases, the relevant scaling factors are the number of registered AASs and submodels, the frequency of operation invocations, the rate of telemetry reads, and the number of active policies that must be translated into P4 table entries. This is different from data-plane scalability, which depends on switch table capacity, rule granularity, and P4Runtime update rate. The current prototype evaluates this mechanism on a small topology, so it validates the control-loop implementation but does not quantify these factors under large-scale concurrency. The IT/control-plane placement of the AAS infrastructure also leaves deployment flexibility. BaSyx registries, dashboards, and historical asset information can be hosted on plant servers, edge platforms, or cloud infrastructure. Components involved in time-sensitive enforcement, such as the Network Manager and the SDN/P4 controller, should remain close to the plant network when low latency or local availability is required. This separation is consistent with the implemented architecture, where AAS services expose asset-level operations while the programmable switches remain responsible for packet forwarding.

In the current prototype, telemetry correlation is implemented through periodic retrieval of network counters and machine-side metrics. This mechanism is appropriate for the evaluated use case, where the number of monitored assets is limited and the sampling interval is fixed. At a larger scale, however, unchanged per-asset polling would increase management traffic and registry/API load. A more scalable implementation should therefore rely on selective monitoring, event-triggered updates, and aggregation of counters at the controller or switch level.

Policy management represents a second scalability dimension. In the prototype, policy updates are serialized by the Network Manager before being forwarded to the SDN controller. This avoids inconsistent updates in the evaluated scenarios. In larger deployments, where multiple AAS instances or applications may request overlapping network actions, the Network Manager should also provide policy arbitration by detecting conflicting traffic selectors and applying deterministic priorities among isolation, maintenance, and ordinary connectivity rules.

7.4. Industry 5.0 Implications

Although the prototype is primarily designed to validate AAS-driven network control in an Industry 4.5 setting, its mechanisms address three core Industry 5.0 dimensions: resilience, human-centricity, and sustainability. These dimensions are not treated as independent add-ons, but as consequences of the same architectural principle: asset-level semantics are exposed through the AAS and translated into programmable network behavior.

The resilience dimension is represented by the closed-loop mitigation mechanism. In the current implementation, the response to a confirmed anomaly is a binary drop rule installed through the SDN/P4 enforcement path. This action provides a deterministic validation of the AAS-to-P4 loop, but the same mechanism can support more selective responses. Since the mitigation is expressed as an AAS-level policy before being translated into data-plane rules, the action can be extended from simple isolation to rate limiting, traffic mirroring, redirection to an inspection path, or time-bounded quarantine. This is relevant for industrial resilience because the objective is not only to stop malicious traffic, but also to preserve the essential communication needed for diagnosis, safety monitoring, and recovery.

https://doi.org/10.3390/fi18070347

The human-centric dimension is illustrated by the maintenance use case. The operator does not manually configure 5G connectivity, OT isolation, and rollback as separate low-level network actions. Instead, the maintenance state is requested through an asset-level AAS operation, and the infrastructure translates this request into the required network configuration. This reduces the operational burden on human operators and makes the resulting network state explicit and auditable. In a production deployment, the same mechanism can be extended with role-based authorization, operator confirmation for disruptive actions, policy expiration times, and rollback conditions. These extensions would preserve human control over safety-relevant decisions while keeping the underlying network orchestration automated.

The sustainability dimension is not directly implemented in the current prototype, but it can be integrated through the same AAS-based abstraction. An energy-aware or environmental submodel could expose information such as energy consumption, carbon intensity, or preferred low-energy connectivity options for network and edge resources.

The prototype, therefore, does not claim to implement a complete Industry 5.0 system. Its contribution is more specific: it demonstrates a programmable and semantic control layer through which Industry 5.0 requirements can be expressed and enforced at the network level. Resilience is addressed through closed-loop mitigation, human-centricity through auditable maintenance operations, and sustainability through the possibility of adding environmental constraints to AAS-based policy decisions.

7.5. Standardization Perspective

The prototype also exposes a standardization gap in current AAS-based industrial modeling. While AAS specifications provide a general structure for representing industrial assets and their submodels, there is no commonly adopted AAS submodel dedicated to programmable network control. In this work, the OT network is therefore modeled through implementation-specific AASs and submodels. This is sufficient to validate the proposed architecture, but it limits portability across different plants, controllers, and network-programming technologies.

Following the SDN principle adopted in the architecture, we argue that network-related AAS standardization should distinguish between control-plane and data-plane functions. A first reusable submodel should describe the Network Control Plane. Its role is to expose the operations and properties needed to express network intent, select or configure controllers, install or remove policies, and track policy state. Typical elements include policy identifiers, target assets, traffic selectors, requested actions, enforcement scope, priority, validity conditions, and installation status. This submodel would represent what the industrial system requires from the network, without binding the AAS interface to a specific enforcement technology.

A second reusable submodel should describe the Network Data Plane. Its role is to expose the observable and enforceable state of the forwarding infrastructure, including switches, ports, counters, flow statistics, table entries, and enforcement points. This submodel would support telemetry retrieval and verification of the rules installed by the control plane. In the prototype, this role is implemented through switch-level functions such as counter retrieval and through P4Runtime-based rule enforcement; in a standardized form, the same abstraction could be mapped to P4, OpenFlow, or other programmable forwarding technologies.

The separation between these two submodels is important because control and data planes evolve differently. The control-plane submodel captures asset-level network intent and policy lifecycle information, whereas the data-plane submodel captures the state and telemetry of the forwarding devices. Keeping these concerns separate makes the

https://doi.org/10.3390/fi18070347

AAS representation more reusable: the same control-plane intent can be translated by the Network Manager into different enforcement backends, while the data-plane submodel can expose technology-specific measurements through a common asset-oriented interface.

This standardization direction is consistent with IEC 63278, which provides the structural basis for AASs and their submodels [16]. The contribution suggested by this work is not a new network protocol, but the definition of two network-oriented AAS submodels: one for the programmable network control plane and one for the programmable network data plane. Such submodels would make asset-aware connectivity, security policies, telemetry, and operational constraints portable across SDN/P4, OpenFlow, 5G, DetNet, and APN-based infrastructures [34–36,64].

8. Conclusions

This paper presented an AAS-driven network control abstraction for converged OT/IT environments. The proposed approach models the OT network as an industrial asset and uses Asset Administration Shells to expose network-control operations, telemetry, and asset-related context through a standardized semantic layer. These AAS-level operations are translated into programmable P4 actions, enabling forwarding policies and mitigation rules to be installed at runtime.

The main contribution is the operational coupling between industrial asset semantics and programmable network enforcement. Rather than treating the network as a passive communication substrate, the proposed architecture makes network behavior controllable through the same asset-oriented abstraction used to represent industrial devices. This coupling enables three capabilities demonstrated in the validation use cases: dynamic customization of forwarding rules, controlled 5G-assisted maintenance with OT isolation and rollback, and anomaly mitigation based on the correlation between AAS-exposed machine and network telemetry.

The evaluation was conducted on a hybrid testbed combining a containerized network environment with a hardware-in-the-loop extension based on Revolution Pi equipment. Across the three use cases, the prototype demonstrates that AAS operations can be used to configure forwarding behavior, coordinate controlled 5G-assisted maintenance, and trigger network-level mitigation from cross-layer telemetry correlation. These results support the feasibility of the proposed abstraction at laboratory scale and justify a TRL 4 assessment, with partial evidence toward TRL 5 for the components exercised with industrial hardware.

Finally, the work identifies a standardization direction for network-oriented AAS models. The implemented architecture separates network control-plane functions from data-plane state and telemetry, following the same separation adopted by SDN-based network design. In this sense, the contribution is not only an implementation of AAS-based network control, but also a first step toward a more portable representation of programmable industrial networks through AAS submodels.

Author Contributions: Conceptualization, R.B., A.M., L.R. and F.C.; methodology, R.B., A.M., L.R. and F.C.; software, R.B.; validation, R.B., A.M. and L.R.; investigation, R.B., A.M. and L.R.; resources, A.M., L.R. and F.C.; data curation, R.B.; writing—original draft preparation, A.M., L.R. and F.C.; writing—review and editing, M.P., R.G., A.M., L.R. and F.C.; visualization, R.B., A.M., L.R. and F.C.; supervision, A.M., R.B., M.P. and F.C.; project administration, A.M., R.B., M.P. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: This study was carried out in the framework of the projects Cri4.0 (CUP: E37G22000490007) and IGNITE5.0 (CUP: E77G22000640003), co-funded by the European funds of the Emilia-Romagna Region, ERDF Regional Programme 2021–2027, Priority 1 Research, Innovation and Competitiveness, Action 1.1.2. This study was also partially funded by the PRIN 2022 Empowering awareness with a digital twin for mission-critical communications (EDIT4MCC)–Codice Progetto 2022Z4N84X_001 CUP J53C24002800006.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

5G	Fifth-Generation mobile network
AAS	Asset Administration Shell
APN	Application-aware Networking
API	Application Programming Interface
CPU	Central Processing Unit
CUPS	Control–User Plane Separation
DetNet	Deterministic Networking
DoS	Denial of Service
DSR	Design Science Research
gRPC	Google Remote Procedure Call
GUI	Graphical User Interface
HIL	Hardware-in-the-Loop
HMI	Human–Machine Interface
HTTP	Hypertext Transfer Protocol
I4.0	Industry 4.0
IAM	Identity and Access Management
ICS	Industrial Control System
IDS	Intrusion Detection System
IEC	International Electrotechnical Commission
IP	Internet Protocol
IPv4	Internet Protocol version 4
IT	Information Technology
LTE	Long-Term Evolution
MITM	Man-in-the-Middle
MQTT	Message Queuing Telemetry Transport
OT	Operational Technology
P4	Programming Protocol-independent Packet Processors
PDU	Protocol Data Unit
PLC	Programmable Logic Controller
QoS	Quality of Service
RAMI 4.0	Reference Architecture Model for Industrie 4.0
RAN	Radio Access Network
REST	Representational State Transfer
RevPi	Revolution Pi
SCADA	Supervisory Control and Data Acquisition
SDN	Software-Defined Networking
TRL	Technology Readiness Level
UE	User Equipment
UI	User Interface

Appendix A. Prototype Screenshots

This appendix provides representative screenshots (Figures A1–A3) of the implemented BaSyx-based prototype, illustrating the AAS interface and the exposed asset information used during the validation scenario.

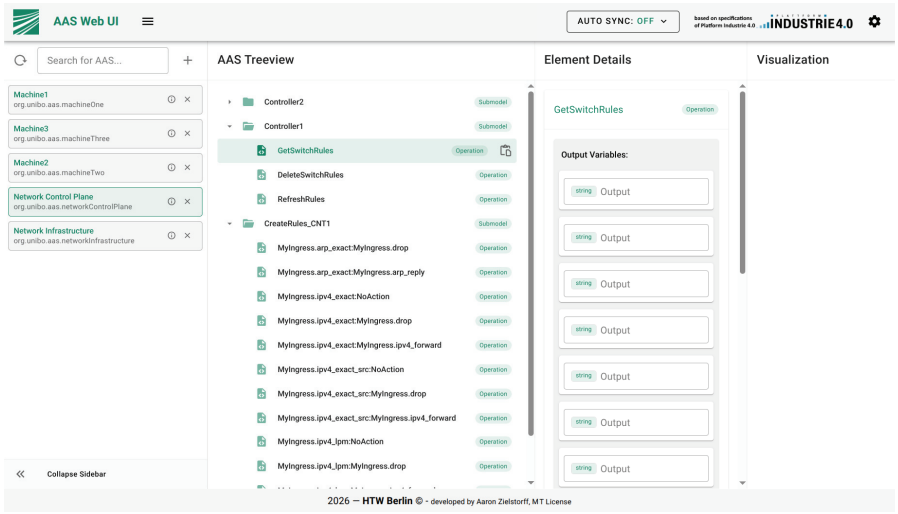


Figure A1. Network control-plane AAS prototype interface.

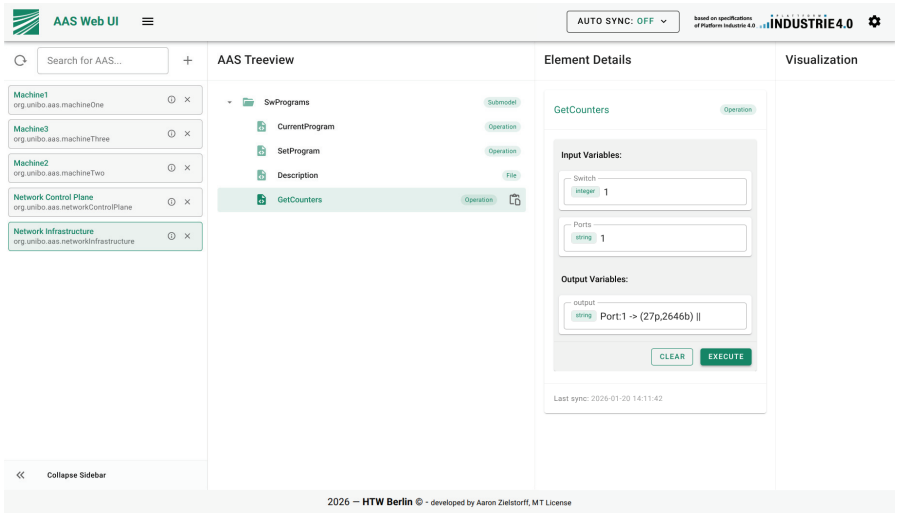


Figure A2. Network data-plane AAS prototype interface.

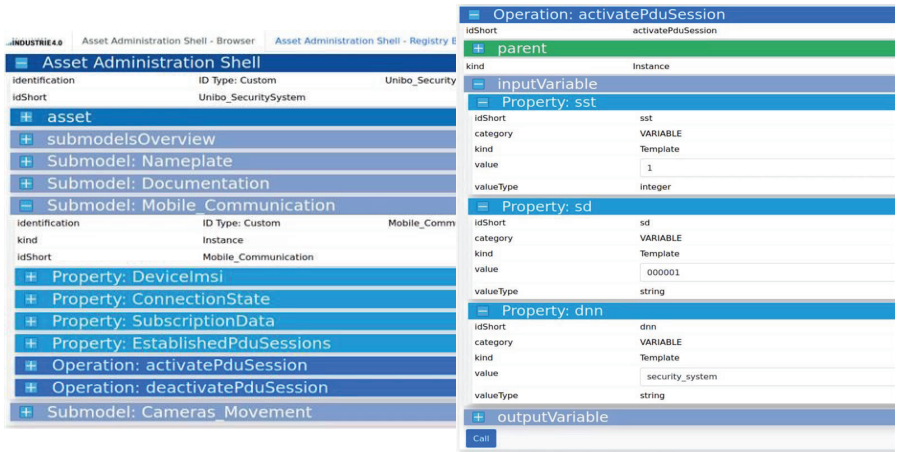


Figure A3. Asset AAS prototype interface with 5G connectivity submodel.

https://doi.org/10.3390/fi18070347

References

1. Berardi, D.; Callegati, F.; Giovine, A.; Melis, A.; Prandini, M.; Rinieri, L. When Operation Technology Meets Information Technology: Challenges and Opportunities. *Future Internet* **2023**, *15*, 95. [CrossRef]
2. Gartner. Operational Technology (OT)—Gartner Glossary. 2022. Available online: https://www.gartner.com/en/information-technology/glossary/operational-technology-ot (accessed on 22 June 2026).
3. Lasi, H.; Fettke, P.; Kemper, H.G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [CrossRef]
4. Ghoson, N.H.; Meyrueis, V.; Benfriha, K.; Guiltat, T.; Loubere, S. A review on the static and dynamic risk assessment methods for OT cybersecurity in industry 4.0. *Comput. Secur.* **2025**, *150*, 104295.
5. Hankel, M.; Rexroth, B. *The Reference Architectural Model Industrie 4.0 (Rami 4.0)*; ZVEI: Frankfurt am Main, Germany, 2015; Volume 2, pp. 4–9.
6. Tantik, E.; Anderl, R. Integrated data model and structure for the Asset Administration Shell in Industrie 4.0. *Procedia Cirp* **2017**, *60*, 86–91. [CrossRef]
7. Wagner, C.; Grothoff, J.; Epple, U.; Drath, R.; Malakuti, S.; Grüner, S.; Hoffmeister, M.; Zimermann, P. The role of the Industry 4.0 asset administration shell and the digital twin during the life cycle of a plant. In *Proceedings of the 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 12–15 September 2017*; IEEE: New York, NY, USA, 2017; pp. 1–8.
8. McKeown, N. Software-defined networking. *INFOCOM Keynote Talk* **2009**, *17*, 30–32.
9. Bosshart, P.; Daly, D.; Gibb, G.; Izzard, M.; McKeown, N.; Rexford, J.; Schlesinger, C.; Talayco, D.; Vahdat, A.; Varghese, G.; et al. P4: Programming protocol-independent packet processors. *ACM SIGCOMM Comput. Commun. Rev.* **2014**, *44*, 87–95. [CrossRef]
10. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *MIS Q.* **2004**, *28*, 75–105. [CrossRef]
11. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A Design Science Research Methodology for Information Systems Research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77. [CrossRef]
12. Plattform Industrie 4.0. *Details of the Asset Administration Shell – Part 1: The Exchange of Information Between Partners in the Value Chain of Industrie 4.0*; Technical Report; Plattform Industrie 4.0: Berlin, Germany, 2020.
13. Lu, Q.; Zhu, D.; Li, M. Intelligent manufacturing with asset administration shell: A bibliometric and knowledge map analysis. *Digit. Eng.* **2026**, *9*, 100077. [CrossRef]
14. Hosseini, A.M.; Sauter, T.; Kastner, W. Towards adding safety and security properties to the Industry 4.0 Asset Administration Shell. In *Proceedings of the 2021 17th IEEE International Conference on Factory Communication Systems (WFCS), Linz, Austria, 9–11 June 2021*; IEEE: New York, NY, USA, 2021; pp. 41–44.
15. Hosseini, A.M.; Sauter, T.; Kastner, W. A safety and security reference architecture for asset administration shell design. In *Proceedings of the 2022 IEEE 18th International Conference on Factory Communication Systems (WFCS), Pavia, Italy, 27–29 April 2022*; IEEE: New York, NY, USA, 2022; pp. 1–8.
16. International Electrotechnical Commission. IEC 63278-1:2023 Asset Administration Shell for Industrial Applications—Part 1: Asset Administration Shell Structure. 2023. Available online: https://webstore.iec.ch/en/publication/65628 (accessed on 28 May 2026).
17. Hauser, F.; Häberle, M.; Schmidt, M.; Menth, M. P4-ipsec: Site-to-site and host-to-site vpn with ipsec in p4-based sdn. *IEEE Access* **2020**, *8*, 139567–139586.
18. AlSabeh, A.; Khoury, J.; Kfoury, E.; Crichigno, J.; Bou-Harb, E. A survey on security applications of P4 programmable switches and a STRIDE-based vulnerability assessment. *Comput. Netw.* **2022**, *207*, 108800. [CrossRef]
19. Pittalà, G.F.; Rinieri, L.; Al Sadi, A.; Davoli, G.; Melis, A.; Prandini, M.; Cerroni, W. Leveraging Data Plane Programmability to enhance service orchestration at the edge: A focus on industrial security. *Comput. Netw.* **2024**, *246*, 110397. [CrossRef]
20. Borsatti, D.; Davoli, G.; Cerroni, W.; Callegati, F. Service Function Chaining Leveraging Segment Routing for 5G Network Slicing. In *Proceedings of the 2019 15th International Conference on Network and Service Management (CNSM), Halifax, NS, Canada, 21–25 October 2019*; IEEE: New York, NY, USA, 2019; pp. 1–6.
21. De Benedictis, A.; Flammini, F.; Mazzocca, N.; Somma, A.; Vitale, F. Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept. *IEEE Trans. Ind. Inform.* **2023**, *19*, 11553–11563. [CrossRef]
22. Varghese, S.A.; Dehlaghi Ghadim, A.; Balador, A.; Alimadadi, Z.; Papadimitratos, P. Digital Twin-based Intrusion Detection for Industrial Control Systems. In *Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops), Virtual Event, 16–18 March 2022*; IEEE: New York, NY, USA, 2022; pp. 611–617. [CrossRef]
23. Alcaraz, C.; López, J. WASAM: A Dynamic Wide-Area Situational Awareness Model for Critical Domains in Smart Grids. *Future Gener. Comput. Syst.* **2014**, *30*, 146–154. [CrossRef]
24. Alem, S.; Espes, D.; Nana, L.; Martin, E.; De Lamotte, F. A novel bi-anomaly-based intrusion detection system approach for industry 4.0. *Future Gener. Comput. Syst.* **2023**, *145*, 267–283.

25. Fung, C.; Srinarasi, S.; Lucas, K.; Phee, H.B.; Bauer, L. Perspectives from a comprehensive evaluation of reconstruction-based anomaly detection in industrial control systems. In *Proceedings of the European Symposium on Research in Computer Security, Copenhagen, Denmark, 26–30 September 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 493–513.
26. Zhang, H.; Xiao, X.; Mercaldo, F.; Ni, S.; Martinelli, F.; Sangaiah, A.K. Classification of ransomware families with machine learning based onN-gram of opcodes. *Future Gener. Comput. Syst.* **2019**, *90*, 211–221.
27. Rossi, D.; Tontini, G.; Borsatti, D.; Callegati, F. Integration of 5G connectivity with the Asset Administration Shell in Industry 4.0. In *Proceedings of the 2022 13th International Conference on Network of the Future (NoF), Ghent, Belgium, 5–7 October 2022*; IEEE: New York, NY, USA, 2022; pp. 1–3. [CrossRef]
28. Perez Olaya, S.S.; Kulasekara Pallewaththe Kankanamge, H.; Cainelli, G.P.; Gambal, B. Management of Industrial 5G Networks over Asset Administration Shell. In *Proceedings of the 2024 IEEE 20th International Conference on Factory Communication Systems (WFCS), Toulouse, France, 17–19 April 2024*; IEEE: New York, NY, USA, 2024; pp. 1–4. [CrossRef]
29. Puligheddu, C.; Ashdown, J.; Chiasserini, C.F.; Restuccia, F. SEM-O-RAN: Semantic O-RAN slicing for mobile edge offloading of computer vision tasks. *IEEE Trans. Mob. Comput.* **2023**, *23*, 7785–7800.
30. Busacca, F.; Faraci, G.; Grasso, C.; Palazzo, S.; Schembra, G. Designing a multi-layer edge-computing platform for energy-efficient and delay-aware offloading in vehicular networks. *Comput. Netw.* **2021**, *198*, 108330.
31. Busacca, F.; Galluccio, L.; Palazzo, S. A marketplace model for drone-assisted edge computing in 5G scenarios. *Comput. Netw.* **2022**, *219*, 109453. [CrossRef]
32. Chen, X.; Zhang, H.; Wu, C.; Mao, S.; Ji, Y.; Bennis, M. Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning. *IEEE Internet Things J.* **2018**, *6*, 4005–4018. [CrossRef]
33. Yu, H.; Wang, H.; Tiwari, D.; Li, J.; Park, S.J. Stellaris: Staleness-Aware Distributed Reinforcement Learning with Serverless Computing. In Proceedings of the SC24: International Conference for High Performance Computing, Networking, Storage and Analysis, Atlanta, GA, USA, 17–22 November 2024.
34. IETF. Application-Aware Networking (APN) Working Group. 2026. IETF Datatracker. Available online: https://datatracker.ietf.org/wg/apn/about/ (accessed on 28 May 2026).
35. IETF Deterministic Networking Working Group. Deterministic Networking (DetNet) Working Group Charter. 2026. Available online: https://datatracker.ietf.org/wg/detnet/charter/ (accessed on 28 May 2026).
36. Finn, N.; Thubert, P.; Varga, B.; Farkas, J. RFC 8655: Deterministic Networking Architecture. IETF RFC 8655. 2019. Available online: https://www.rfc-editor.org/info/rfc8655/ (accessed on 22 June 2026).
37. Graessler, I.; Pöhler, A. Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system. In *Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*; IEEE: New York, NY, USA, 2017; pp. 289–293.
38. Cheng, Z. Human digital twin with applications. In Proceedings of the 7th International Digital Human Modeling Symposium (DHM 2022) and Iowa Virtual Human Summit 2022, Iowa City, IA, USA, 29–31 August 2022.
39. International Telecommunication Union. Recommendation ITU-T L.1333: Carbon Data Intensity for Network Energy Performance Monitoring. 2022. Available online: https://www.itu.int/rec/T-REC-L.1333-202209-I (accessed on 28 May 2026).
40. International Telecommunication Union. Recommendation ITU-T L.1470: Greenhouse Gas Emissions Trajectories for the Information and Communication Technology Sector Compatible with the UNFCCC Paris Agreement. 2020. Available online: https://www.itu.int/ITU-T/recommendations/rec.aspx?id=14084 (accessed on 28 May 2026).
41. International Telecommunication Union. Recommendation ITU-T L.1480: Enabling the Net Zero Transition–Assessing How the Use of Information and Communication Technology Solutions Impacts Greenhouse Gas Emissions of Other Sectors. Recommendation ITU-T L.1480 (07/2025), in Force. 2025. Available online: https://www.itu.int/rec/T-REC-L.1480-202507-I (accessed on 28 May 2026).
42. International Telecommunication Union. ITU-T Study Group 5: Environment, Climate Action, Circular Economy and Electromagnetic Fields. 2026. Available online: https://www.itu.int/en/ITU-T/studygroups/2025-2028/05/Pages/default.aspx (accessed on 28 May 2026).
43. Patzer, F.; Volz, F.; Usländer, T.; Blöcher, I.; Beyerer, J. The industrie 4.0 asset administration shell as information source for security analysis. In *Proceedings of the 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 10–13 September 2019*; IEEE: New York, NY, USA, 2019; pp. 420–427.
44. Rose, S.; Borchert, O.; Mitchell, S.; Connelly, S. Zero Trust Architecture. Special Publication 800-207, National Institute of Standards and Technology, 2020. Available online: https://csrc.nist.gov/pubs/sp/800/207/final (accessed on 22 June 2026).
45. Williams, T.J. The Purdue enterprise reference architecture. *Comput. Ind.* **1994**, *24*, 141–158. [CrossRef]
46. Dolev, D.; Yao, A. On the security of public key protocols. *IEEE Trans. Inf. Theory* **1983**, *29*, 198–208. [CrossRef]
47. Abbas, S.G.; Ozmen, M.O.; Alsaheel, A.; Khan, A.; Celik, Z.B.; Xu, D. SAIN: Improving ICS Attack Detection Sensitivity via State-Aware Invariants. In *Proceedings of the 33rd USENIX Security Symposium (USENIX Security 24), Philadelphia, PA, USA, 11–13 August 2024*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 6597–6613.

48. Ike, M.; Phan, K.; Sadoski, K.; Valme, R.; Lee, W. Scaphy: Detecting modern ics attacks by correlating behaviors in scada and physical. In *Proceedings of the 2023 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–25 May 2023*; IEEE: New York, NY, USA, 2023; pp. 20–37.
49. Pickren, R.; Shekari, T.; Zonouz, S.; Beyah, R. Compromising industrial processes using web-based programmable logic controller malware. In Proceedings of the Network and Distributed System Security (NDSS) Symposium, San Diego, CA, USA, 26 February–1 March 2024.
50. Erba, A.; Tippenhauer, N.O. Assessing model-free anomaly detection in industrial control systems against generic concealment attacks. In *Proceedings of the 38th Annual Computer Security Applications Conference, Austin, TX, USA, 5–9 December 2022*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 412–426.
51. Industrie4.0. Basyx Explained. Available online: https://wiki.basyx.org/en/latest/content/introduction/basyx_explained.html (accessed on 22 June 2026).
52. 3GPP. 3GPP TS 23.214: Architecture Enhancements for Control and User Plane Separation of EPC Nodes (Release 19). 2025. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3077 (accessed on 22 June 2026).
53. 3GPP. 3GPP TS 23.401: General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access (Release 19). 2025. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=849 (accessed on 22 June 2026).
54. 3GPP. 3GPP TS 23.501: System Architecture for the 5G System (5GS) (Release 19). 2025. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144 (accessed on 22 June 2026).
55. Liatifis, A.; Sarigiannidis, P.; Argyriou, V.; Lagkas, T. Advancing SDN from OpenFlow to P4: A Survey. *ACM Comput. Surv.* **2023**, *55*, 1–37. [CrossRef]
56. Bonofiglio, G.; Iovinella, V.; Lospoto, G.; Di Battista, G. Kathará: A container-based framework for implementing network function virtualization and software defined networks. In Proceedings of the *NOMS 2018—2018 IEEE/IFIP Network Operations and Management Symposium, Taipei, China, 23–27 April 2018*; IEEE: New York, NY, USA, 2018; pp. 1–9.
57. The P4 Language Consortium. P4Runtime Specification, Version 1.5.0. 2026. Available online: https://p4lang.github.io/p4runtime/spec/v1.5.0/P4Runtime-Spec.html (accessed on 22 June 2026).
58. Aheleroff, S.; Xu, X.; Zhong, R.Y.; Lu, Y. Digital Twin as a Service (DTaaS) in Industry 4.0: An Architecture Reference Model. *Adv. Eng. Inform.* **2021**, *47*, 101225. [CrossRef]
59. Liu, M.; Fang, S.; Dong, H.; Xu, C. Review of Digital Twin About Concepts, Technologies, and Industrial Applications. *J. Manuf. Syst.* **2021**, *58*, 346–361. [CrossRef]
60. Xu, H.; Wu, J.; Pan, Q.; Guan, X.; Guizani, M. A survey on digital twin for industrial internet of things: Applications, technologies and tools. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 2569–2598. [CrossRef]
61. Munir, M.; Siddiqui, S.A.; Chattha, M.A.; Dengel, A.; Ahmed, S. FuseAD: Unsupervised anomaly detection in streaming sensors data by fusing statistical and deep learning models. *Sensors* **2019**, *19*, 2451. [CrossRef] [PubMed]
62. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, K.R. A Unifying Review of Deep and Shallow Anomaly Detection. *Proc. IEEE* **2021**, *109*, 756–795. [CrossRef]
63. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation Forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008*; IEEE: New York, NY, USA, 2008; pp. 413–422. [CrossRef]
64. International Telecommunication Union. ITU-T Study Group 13: Future Networks and Emerging Network Technologies. Study Period 2025–2028. 2026. Available online: https://www.itu.int/en/ITU-T/studygroups/2025-2028/13/Pages/default.aspx (accessed on 28 May 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/fi18070347

future internet

MDPI

Article

Smart IoT-Based Temperature-Sensing Device for Energy-Efficient Glass Window Monitoring

Vaclav Mach, Jiri Vojtesek *, Milan Adamek, Pavel Drabek, Pavel Stoklasek, Stepan Dlabaja, Lukas Kopecek and Ales Mizera

Faculty of Applied Informatics, Tomas Bata University in Zlin, Nad Stranemi 4511, 760 05 Zlin, Czech Republic; v2mach@utb.cz (V.M.); adamek@utb.cz (M.A.); pdrabek@utb.cz (P.D.); pstoklasek@utb.cz (P.S.); sdlabaja@utb.cz (S.D.); l_kopecek@utb.cz (L.K.); mizera@utb.cz (A.M.)
* Correspondence: vojtesek@utb.cz

Abstract

This paper presents the development and validation of an IoT-enabled temperature-sensing device for real-time monitoring of the thermal insulation properties of glass windows. The system integrates contact and non-contact temperature sensors into a compact PCB platform equipped with WiFi connectivity, enabling seamless integration into smart home and building management frameworks. By continuously assessing window insulation performance, the device addresses the challenge of energy loss in buildings, where glazing efficiency often degrades over time. The collected data can be transmitted to cloud-based services or local IoT infrastructures, allowing for advanced analytics, remote access, and adaptive control of heating, ventilation, and air-conditioning (HVAC) systems. Experimental results demonstrate the accuracy and reliability of the proposed system, confirming its potential to contribute to energy conservation and sustainable living practices. Beyond energy efficiency, the device provides a scalable approach to environmental monitoring within the broader future internet ecosystem, supporting the evolution of intelligent, connected, and human-centered living environments.

Keywords: temperature-sensing device; internet of things; thermal insulation; window

1. Introduction

Temperature-sensing devices have been widely employed in the fields of medicine, industry, agriculture, and other areas and have become the principal method of information collection due to the rapid development of microelectronics and sensor technology. These sensors are placed at key locations within the monitored area to capture thermal characteristics with high accuracy, and then are usually transmitted to users via wireless communication equipment [1,2]. Distributed temperature sensors with high processing capability and low power consumption are used in smart homes, and they can capture data on a resident's interactions with the environment [3]. These settings are designed to recognize situations to support users in their daily activities and offer smart solutions to some of the issues brought on by the expanding population [4].

These devices can be utilized to lower the energy consumption of building components that may lose their insulating capacity over time. The most vulnerable component of the building is the window where the outside protective structure's heat insulation is the weakest. Additionally, it is a crucial tool for visual performance, which contributes significantly to building energy savings. Contradictions exist between energy consumption,

visual performance, and thermal environment because of the weather and sun radiation. Consequently, an efficient optimization technique is required to optimize the mentioned parameters. Most of the studies that are now available for window design concentrate more on the examination of energy consumption performance. Additionally, the best way to arrange windows and a shading system with various climatic zones and orientations has been determined. The building's energy use has increased significantly in recent years. A recent study [5] stated that 36% of the world's energy usage and carbon emissions have been attributed to it. Due to a lack of energy and high energy prices, the modern building industry is paying increasing attention to energy efficiency. Countries have implemented relevant energy-saving regulations and policies, and energy efficiency as an aspect of architectural design is a crucial factor [6].

Moreover, purpose design can be seamlessly integrated into smart home systems by IoT technologies, enabling additional functionalities such as heating control, energy management, and other smart home applications [7]. This multifunctional approach enhances the overall utility of such systems, making them adaptable to various user needs. To achieve this level of functionality, researchers frequently employ platforms like Arduino or Raspberry Pi, which provide the computational power and flexibility required to meet the system's demands [8,9]. These platforms enable the creation of advanced, integrated devices that combine energy efficiency with smart home capabilities.

Multi-Layer Glazing

Nowadays, most of the windows in private or industrial areas are made of multi-layer glazing, which has excellent thermal insulation. Moreover, the space between the two glass panels is filled with special gas to reduce the thermal conductivity between these glass panels [10]. Most commonly used internal gases in multi-layer glazing are argon (Ar) and krypton (Kr), which have better insulating properties than ordinary air [11].

Contemporary insulators include vacuum insulation panels [12], gas-filled panels, aerogels [13], and thermal insulation materials. The primary characteristics of insulating materials include very low thermal conductivity and their potential transparency. The authors of [14,15] point out that the gas will leak over time. The rate of leakage depends on the production quality of the window, installation, the building's climate, including its exposure to the sun and altitude, and other factors. Even after several years of gradual depressurization, a window can be expected to maintain its properties and effectiveness. The stress between the glass panels and the seal due to cyclic thermal loading causes the deformation and weakening of the seal, which leads to the infiltration of gases into the glass chamber. This can lead to the dilution of the inert gas concentration [16].

There are two key components of the multi-layer glazing construction. The first one consists of glass panels that are attached together by a frame and special glue. A special gas is used to fill the space between the glass panels, reducing the thermal conductivity of the area between two or more glass panels [17]. For example, argon gas is heavier than oxygen and has less energy and heat transmission than diatomic molecules. The second component is the plastic frame that holds the glass panel. Typically, the structure is separated into multiple chambers to increase insulation properties [18].

The gas leakage or the deterioration of the inner gas properties inside the multi-layer glazing is very difficult to measure or evaluate. One possible way to measure the gas properties between two glass panels is to place a sensor inside the gap during the assembly. Another way is to place the needed sensor inside the multi-layer glazing, where it can measure environmental variables, as well as the temperature.

This study presents a compact environmental monitoring system for multi-layer glazing, capable of measuring both contact and gas temperatures. The versatile temperature-

sensing device supports various IoT and smart home applications, enhancing energy efficiency and offering valuable insights into long-term insulation property trends.

2. Related Work

Several authors have addressed heat transfer in multi-layer glazing systems using thermocouples. The authors of [19] used several thermocouples to measure the temperature inside triple-chamber glazing. Two thermocouples were placed inside each cavity to measure the heat transfer through the glazing. However, this research focuses on the comparison of the thermal performance of four triple-glazed window configurations (TG, TG-LW, TG-PCM, TG-LW-PCM) in the range from 20 °C up to 50 °C. Another author [20] used a similar approach by using the TRNSYS program within a temperature range from 17 °C up to 40 °C with an additional cooling phase. The authors of [21] used different simulation models for glazing windows filled with a special phase change material. These authors have stated that accurate thermal monitoring within the glazing system requires measuring the glass panel's contact temperature and the air temperature inside the cavity.

A practical approach is to use a compact sensor module mounted on a small PCB. Thermocouples or resistance temperature detectors are ideal due to their minimal thermal mass and fast response time for contact measurement of the temperature. Digital sensors like BMP280 offer high accuracy and low power consumption for temperature measurement [22]. Moreover, the sensor is very small with dimensions of $2.5 \times 2.5 \times 1$ mm [23], and it is suitable for temperature measurement, which is used as a basis task for this research.

Furthermore, the authors of [24] proposed integrating ultra-thin flexible temperature sensors laminated directly onto the glass surface to minimize disturbance of the thermal field while enabling high-resolution spatiotemporal monitoring. Other research [25] has focused on wireless passive sensors, such as RFID-based temperature probes, which can be hermetically sealed within the glazing cavity during manufacture. Furthermore, machine-learning-assisted interpretation of temperature profiles has been investigated as a method for estimating the remaining insulating lifespan of glazing systems under varying environmental loads [26]. These solutions demonstrate the growing interest in non-intrusive diagnostic tools capable of detecting changes in thermal properties.

3. Materials and Methods

This section outlines the materials, experimental setup, and analytical methods employed to evaluate the thermal and environmental performance of the proposed system. The research integrates advanced simulation techniques with controlled physical experiments to provide a comprehensive understanding of the heat transfer dynamics, sensor functionality, and overall energy efficiency of the window.

In common applications, such a window is intended to have dimensions of up to several meters, reflecting the standard sizes commonly used in building construction for energy-efficient glazing systems. This full-scale design would allow for a comprehensive evaluation of the window's performance in practical scenarios, including its thermal insulation capabilities and integration with building management systems. However, for this research, a scaled-down model with dimensions of 200×200 mm was designed. This smaller prototype was designed to replicate the essential features of the larger window while remaining applicable for computational simulations and laboratory experiments. The reduced size enabled temperature control over the experimental conditions and facilitated the efficient application of computational fluid dynamics (CFD) techniques, which require detailed meshing and significant computational resources.

The designed system in the form of several printed circuit boards (PCBs) (Printed s.r.o., Melnik, Czech Republic) with sensors is placed inside the glazing on the inner and outer glass panels. Two boards were placed on each glass panel, where all boards consisted of the microcontroller with a wireless communication interface and a pair of BME280 environmental sensors (Printed s.r.o., Melnik, Czech Republic). All boards were glued to the side of each glass panel for temperature measurement. Each sensor was strategically positioned on the board, ensuring accurate readings of both indoor and outdoor environmental conditions. The sensors operate within a wide temperature range, making them suitable for measuring environmental conditions. The concept can be extended to any number of chambers, where each chamber needs to have four boards. The proposed concept is shown in Figure 1.

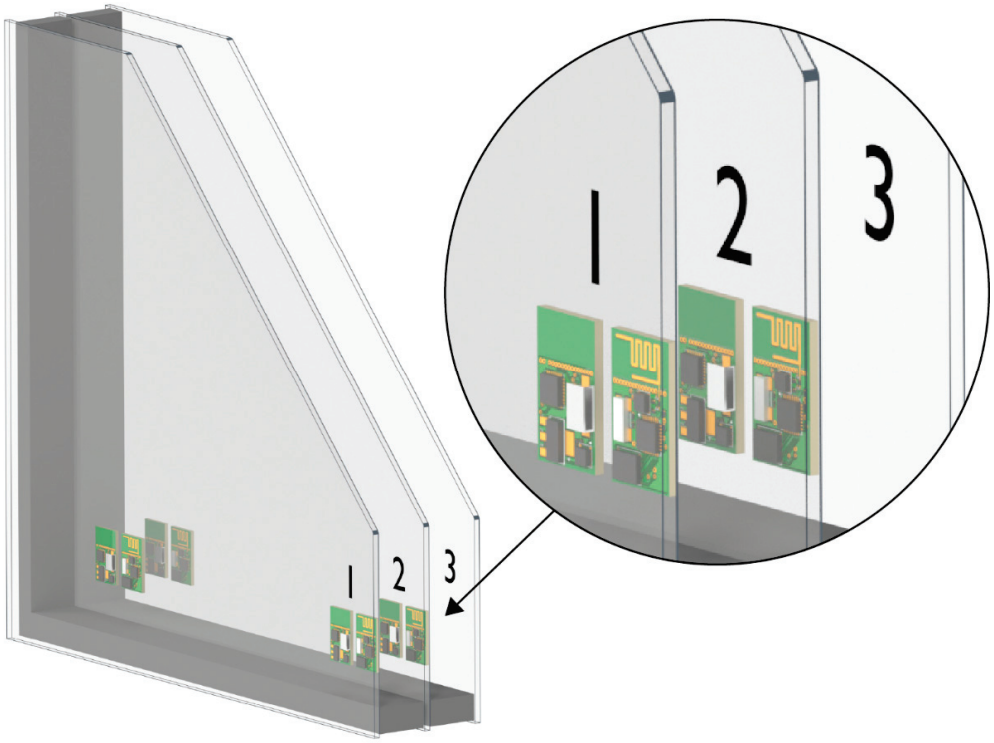


Figure 1. The concept of temperature-sensing device design placed in multi-layer glazing.

The proposed system was designed to seamlessly integrate with modern building management systems. It can provide real-time environmental data to optimize heating, ventilation, and air-conditioning (HVAC) operations, reducing energy consumption and improving. In this research, a compact model facilitated the setup of controlled thermal gradients and sensor measurements, ensuring accurate data collection without the logistical challenges associated with handling a full-sized window. Despite its smaller dimensions, the prototype maintained the key characteristics of the larger window, including gas-filled gaps and integrated sensors for temperature. By scaling down the window for this study, it was possible to test its performance while ensuring that the results could be extrapolated to a full-scale implementation. This approach provided a practical and efficient pathway to validate the system's functionality and establish its potential for energy efficiency applications. The scaled-down model for both scenarios can be found in Figure 2.

The chamber system incorporates standard glass panels with gas-filled gaps to enhance thermal insulation. Each panel was made from low-emissivity glass, which reduces heat transfer and improves energy efficiency. The inner and outer panels are coated with a thin metallic layer to reflect infrared radiation, while the middle panels are left untreated to facilitate temperature distribution analysis. The frame between the panels was constructed

from aluminum, and the sealing material ensures airtightness to maintain consistent environmental conditions within the gaps. The list of material properties of the single-layer glazing is listed in Table 1.

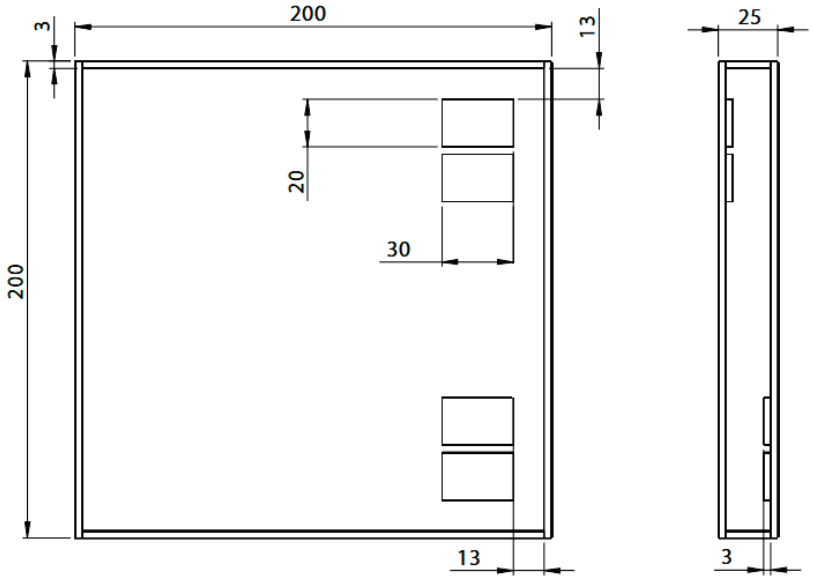


Figure 2. The scaled-down model of the single-layer glazing.

Table 1. Material properties of the single-layer glazing.

Material	Density [kg/m]	Specific Heat Capacity [J/(kg·K)]	Thermal Conductivity [W/(m·K)]
Air	1.2	1005.0	0.1
Glass-low-e	2464.9	898.6	0.9
Semiconductor	3098.4	658.0	117.4
Copper	8978.0	381.0	387.6
Industrial carbon	1529.7	959.8	12.2
Aluminum	2719.0	871.0	202.4

3.1. Setup for the Real Experiment

In the experimental setup, each embedded ESP8266-based sensing board (Printed s.r.o., Melnik, Czech Republic) communicated with a nearby ESP8266 access point positioned 1 m from the glazing surface. Although the PCB was installed inside the window cavity, the short transmission distance ensured stable WiFi connectivity throughout the entire measurement period. While full RF characterization (RSSI, packet loss rate, or latency under varying thermal conditions) was not performed, the stable operation at this range is consistent with typical ESP8266 performance, where RSSI values around -40 to -55 dBm are expected at a line-of-sight distance of 1 m, even with moderate attenuation from glass. For data transmission, each embedded board operated as an independent client node. Outside the glass panel, an additional ESP8266 module was configured as a WiFi access point (AP) acting as the central server. During the experiment, all boards established a wireless connection to this AP and periodically transmitted their measured temperature values. The access point was connected to a PC, where all incoming data were collected and stored for further analysis. This architecture enabled reliable communication between the sensor boards and the data acquisition system without requiring any physical data cables within the window chamber (Figure 3). No data retransmissions or communication interruptions were observed during the experiment, indicating that the system remained sufficiently reliable for continuous temperature reporting over this short-range link.

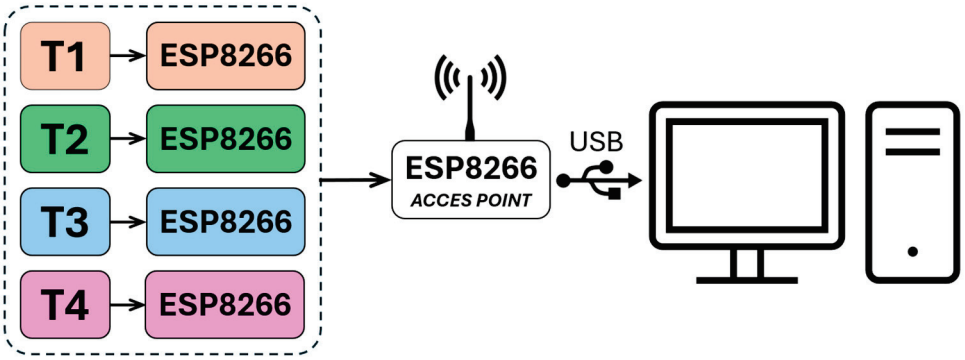


Figure 3. Communication architecture of the proposed system.

3.2. Temperature Sensors for Glazing Applications

The core of the system is based on the ESP8266 microcontroller, chosen for its reliability and flexibility in handling sensor data and wireless communication. It supports low-power operation to align with the energy efficiency goals of the smart window. The microcontroller's firmware is programmed to process sensor data, namely the temperature received from the connected sensors. The schematic part of the board can be found in Figure 4.

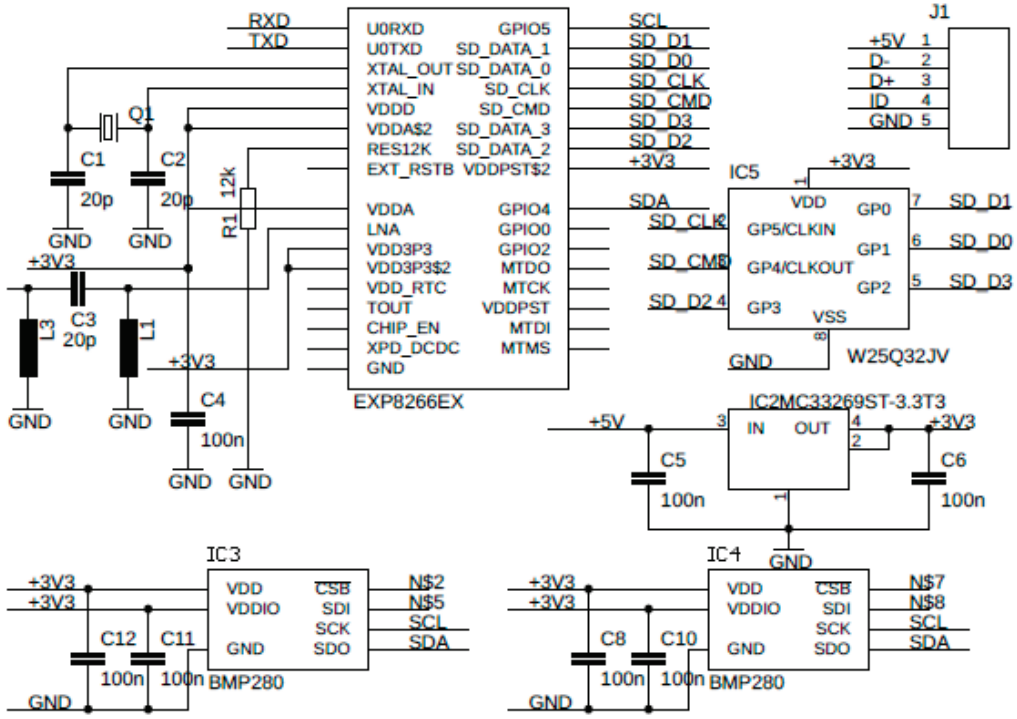


Figure 4. The schematic design of the board embedded into a single −layer glazing.

The board has all the components necessary for external and internal communication. The main microcontroller placed on the board, ESP8266, is responsible for internal communication with sensors, and it is also responsible for external wireless communication. It uses the 40 MHz crystal for the lower power voltage of 3.3 V. All components use the I2C communication interface, which is used for communication between the sensors and the main microcontroller. The main microcontroller has a small package, an important communication interface, and sufficient computing power.

The input voltage of +5V for the whole device is supplied by the USB terminal, and it must be converted to +3V3, which is suitable for the different components in the design. It also has the input and output sets of the capacitors, which can smooth the output voltage

in case of a sinusoidal input. The final physical design of the PCB must be small enough to fit in the gap inside the window frame. In the study by Osorio-de-la-Rosa et al. [27], a comparable compact IoT sensing platform is presented, where the complete wireless sensor is implemented on a 20 × 20 mm PCB. This provides a useful engineering benchmark for the required miniaturization of embedded sensing electronics. In our case, although the ESP8266 module imposes slightly larger spatial requirements, the resulting PCB layout of 20 × 30 mm remains within the dimensional limits of the glazing cavity and satisfies the constraints for integration inside the window chamber. The terminal for the USB interface is placed at the edge of the mainboard. The Wi-Fi antenna is located on the top part of the PCB.

3.3. Manufacturing Process of the Single-Layer Glazing

The designed boards were manufactured from a composite material FR-4 1,5/Cu 35 µm, which is a usual material and can be manufactured by a common manufacturer on the market. The soldering of the components on the board is also a standard procedure. For this research, the components were soldered on the board by an external manufacturer. Figure 5 presents a sectional view of a single-glazed window, illustrating the placement of four temperature sensors (T1–T4) used for thermal analysis.

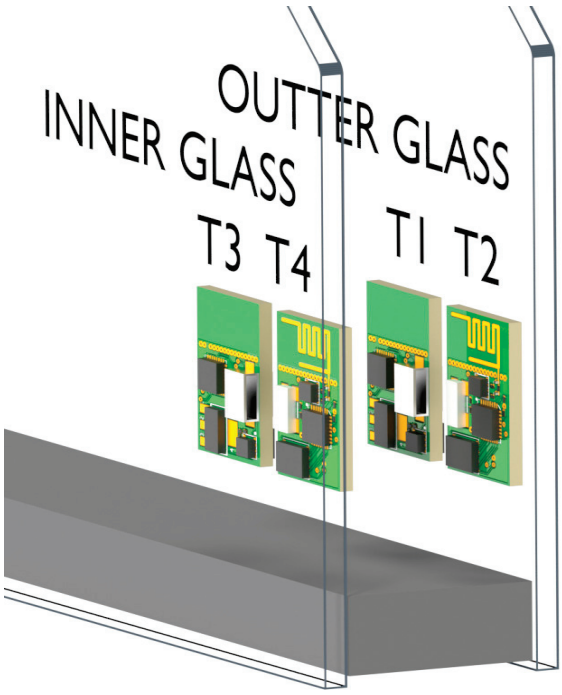


Figure 5. The placement of the sensors inside the single-layer glazing.

Two boards were positioned on the outer glass panel: T1, which measured the temperature of the outer glass surface, and T2, which measured the temperature of the environment close to the outer glass. On the inner glass panel, T3 measured the environment temperature near the inner glass, while T4 measured the temperature of the inner glass surface. This setup enabled a detailed evaluation of heat transfer across the window, capturing temperature variations between the inner and outer environments and assessing the insulating performance of the gas between the glass panels.

All boards with sensors and other electronic components were embedded into the chamber of the window in the fabrication process, where the boards were attached to the glass panels. A heat-conducting adhesive was used for attaching the boards to the glass panels. For the experiments, the boards were powered using an ultra-thin flexible PCB

cable. Its minimal thickness ensured that it did not create a significant thermal bridge inside the window chamber, allowing the temperature measurements to remain unaffected by heat conduction from the power supply. Once the boards were glued to the glass panel and connected via the flexible strip, the aluminum frame could be placed along the edges of the glass panel. Finally, the second glass panel could be placed on the frame with the boards closed inside. The same manufacturing process can also be used for multi-layer glazing.

3.4. Experimental and Simulation Approaches

To comprehensively evaluate the thermal insulation properties of the developed temperature-sensing device, two complementary approaches were employed: computational modeling and laboratory testing. First, a Computational Fluid Dynamics (CFD) simulation was carried out to analyze heat transfer mechanisms and temperature distribution across the single-glass window under controlled boundary conditions. This allowed for a detailed examination of thermal behavior without external disturbances. Second, a real experimental setup was designed to validate the simulation results under practical conditions, using heated air and integrated sensors to measure thermal responses of the glazing system. Together, these approaches provided both predictive insights and empirical verification, ensuring robust assessment of the device's performance.

3.4.1. Solution Approach Using CFD Simulation

The performance of the smart window system was evaluated through CFD simulation. The window was designed as a single-glazed structure equipped with integrated sensors to monitor the surface temperature of the environment inside the window. These sensors were strategically placed to capture accurate data across the window. The primary goal of the analysis was to investigate the heat transfer characteristics across the single-layered construction of the window. Additionally, the study aimed to determine the temperature distribution at specific measurement points, including the inner and outer surfaces of the glass layers and the air cavities in between.

The operating conditions were selected to simulate realistic environments. The outdoor air temperature range was set from 0 °C up to 73 °C, while the indoor air temperature was set to 15 °C, reflecting an indoor environment. These temperatures created a gradient, enabling the assessment of thermal performance under challenging conditions. This approach ensured that the results were relevant to real scenarios and could contribute to the development of energy-efficient glazing systems.

The CFD simulation was designed to achieve a detailed analysis of the thermal performance of the window system. To provide a more accurate representation of the thermal environment, two distinct regions were created within the model. The inner region represented the indoor space adjacent to the window, while the outer region represented the outdoor space exposed to external environmental conditions. In both regions, low-intensity airflow was assumed to mimic natural convection effects. The equation that represents the conservation of energy in a physical system and is crucial for understanding heat transfer is listed in Equation (1).

$$\frac{\partial}{\partial t}\left(\rho\left(e+\frac{v^2}{2}\right)\right)+\nabla\cdot\left(\rho v\left(h+\frac{v^2}{2}\right)\right)=\nabla\cdot\left(k_{eff}\nabla T-\sum_{j}^{1}h_j\vec{J}_j+\bar{\bar{\tau}}_{eff}\cdot\vec{v}\right)+S_h \quad (1)$$

Equation (1) balances the total energy within a system by accounting for energy storage, transport, and sources. This balance is vital in applications such as thermal analysis of materials, fluid dynamics, and the design of systems such as energy-efficient windows.

The left-hand side of the equation expresses energy storage and transport. The first term denotes the rate of change of total energy (internal plus kinetic) within the system,

showing how energy is accumulated or depleted over time. The second term represents convective transport, where fluid motion carries energy in the form of enthalpy and kinetic energy. Together, these terms describe how energy evolves and moves through the system.

The right-hand side accounts for mechanisms of energy transfer and generation. The first term corresponds to heat conduction, governed by Fourier's law, which drives heat flow through temperature gradients. The second term captures energy transport due to mass diffusion, as different species in a mixture contribute to the balance. The third term reflects viscous dissipation, where internal friction converts mechanical energy into heat. Finally, the volumetric source term includes additional energy inputs or losses, such as those from radiation, chemical reactions, or electrical processes.

To investigate the heat transfer and thermal insulation performance of the proposed smart window system, a comprehensive computational fluid dynamics (CFD) simulation was conducted, closely aligned with the experimental configuration. The digital model was constructed to replicate the physical test sample in all key geometric and material aspects, ensuring that the simulation results would be directly comparable to empirical data.

The CAD model defined all relevant physical domains: solid components, such as the glass panes, the structural aluminum frame, and the printed circuit boards (PCBs) housing the sensors, as well as multiple fluid zones representing the interior room air, the external environment, and the insulating cavity between the glass layers. The spatial placement and thickness of each domain strictly followed the actual test specimen, with precise accommodation of the sensor boards' geometry, since their thermal interaction with the glazing system may influence local heat flow and temperature gradients.

The computational mesh was generated using Ansys Fluent (version 2024R1), leveraging polyhedral elements to achieve greater flexibility and accuracy in modeling complex geometries and fine structures. The total mesh density reached 13.3 million cells, with targeted refinement in regions of expected high thermal gradients and velocity shear, notably within the insulating cavity and near the sensors. Throughout the mesh generation process, particular attention was paid to the wall-adjacent flow zones, reflected in adaptive local refinements based on the evolving wall function parameter Y+, which was carefully constrained to remain below 1. Whenever steeper gradients developed during the simulation, automated mesh adaptation further improved local resolution. Mesh-independence was assessed through simulations on several mesh densities; when increasing the mesh density led to negligible variation (below 2%) in the tracked temperatures and heat flows, the grid was deemed sufficiently converged for reliable results.

The CFD solver setup was based on a pressure-based, pseudo-transient approach, utilizing a relaxation factor of 0.5 to ensure numerical stability while capturing the thermal dynamics efficiently. Pressure–velocity coupling employed the pseudo-transient scheme, with momentum and energy equations discretized at second-order upwind accuracy, and pressure discretization set to the body-force-weighted scheme. Radiative transfer was included using the Discrete Ordinates (DO) model, discretized at first order upwind. These choices guaranteed sufficient accuracy for both strong transient behavior and smooth propagation of heat fronts through solid and fluid regions.

A detailed analysis of the flow regime was conducted prior to simulation via computation of dimensionless numbers—Prandtl, Grashof, and Reynolds—for each fluid domain. While laminar conditions prevailed in some smaller cavities, most regions experienced transitional or low-turbulence flow due to the imposed thermal gradients and geometric constraints. Accordingly, the k-ω SST turbulence model was selected for its well-documented ability to handle both laminar-to-turbulent transitions and near-wall effects accurately, ensuring robust simulation of convective heat transfer.

Material properties were defined with high fidelity: air was modeled as an incompressible ideal gas, with temperature-dependent density (using the NASA 9-piecewise polynomial function for specific heat) and thermal conductivity (calculated by kinetic theory), while dynamic viscosity followed Sutherland's law. Solid domains, including glass, aluminum, and PCBs, were assigned literature-based values for density, specific heat capacity, and thermal conductivity. Of particular importance is that the thermal contact resistance between the sensor PCB and the inner glass surface was physically modeled as a thin resistive layer, corresponding to the heat-conducting adhesive used in the assembly, with resistance values matching manufacturer data (~0.0015 m^2K/W), a factor shown to be critical in local thermal dynamics.

Boundary and initial conditions were rigorously derived from the experimental protocol. The window system's outer surface was exposed to a controlled heating ramp, starting at 3 °C and rising to 73 °C according to the linear formula theat = 0.031 [°C s^{-1}] * t + 3 [°C]. Ambient temperature was maintained at 11 °C for both interior and exterior domains. Gravity acted in the negative y-direction, instigating natural convection within the air cavities and ensuring that buoyancy-driven flows reproduced their real-life counterparts. All initial temperatures within fluid and solid domains matched their respective ambient conditions.

Time-stepping was assigned with great care to ensure numerical robustness. The minimum characteristic length scales-primarily the thicknesses of the glass and PCB-dictated a strict maximum stable time-step via the CFL criterion, found to be approximately 3 s. The simulation adopted an adaptive approach; initial highly transient phases were resolved with a time-step of 0.005 s to accurately capture rapid changes, after which the step size was gradually increased to 0.5 s as gradients stabilized. For every time step, 20 solver iterations were performed, and key residuals were tightly controlled: energy equations required convergence below (10^{-6}) to (10^{-7}), while other equations converged below (10^{-4}) to (10^{-5}). Additionally, heat flow and temperature changes were monitored to ensure solution change below 1% between iterations.

Throughout the simulation, critical quantities were continuously monitored—sensor temperatures, maximum and average temperatures across each component, heat fluxes through the glazing system, and maximum flow velocities within the fluid domains. These data streams enabled both local and global validation against experimental measurements and facilitated multipoint comparison of temperature trajectories over time. The entire post-processing procedure was carried out using Ansys Fluent's suite of visualization and analysis tools.

This sophisticated computational approach, with the rigorous definition of all mesh, solver, boundary, and material parameters, guarantees methodological transparency and reproducibility. By capturing and quantitatively evaluating heat transfer phenomena at high spatial and temporal acuity, the CFD simulation provides a solid foundation for direct comparison with experiment and ensures the scientific credibility of all subsequent claims regarding the insulation performance and dynamic behavior of the smart window system.

3.4.2. Solution Approach Using Real Experiment

The experimental setup was designed to closely replicate the boundary conditions of the computational model. Hot air was used as the heat transfer medium. An electric oven was used to uniformly heat the outer glass panel from 0 to 73 °C. The manufactured boards with temperature sensors were placed on the glass surfaces within each chamber and secured with heat-conductive adhesive to ensure accurate data collection.

The experiment was conducted in a controlled laboratory environment at a constant temperature of 15 °C and 50% relative humidity, eliminating external variables such as ambient temperature fluctuations. The air temperature was monitored with a digital

thermometer and heated incrementally to minimize thermal shock and accurately observe the glazing's response. This controlled setup ensured reliable results focused solely on the interaction between the glazing and the heated air.

3.5. Temperature Sensor Calibration

The temperature sensors BME280 used in the prototype were calibrated prior to the experiment using a traceable reference thermometer (accuracy ±0.1 °C, ISO/IEC 17025 [28] certified). The calibration was performed inside a temperature-controlled chamber in the range of 10–80 °C, which corresponds to the expected operating conditions of insulated glazing units. A total of eight calibration points (from 10 °C to 80 °C) were selected with a dwell time of 10 min at each temperature to ensure thermal stabilization.

At every calibration point, five repeated measurements were taken to evaluate sensor repeatability, for which the coefficient of variation remained below 0.2%. A linear regression model was used to fit the measured values to the reference instrument, as no significant nonlinearity was detected within the calibration range. The resulting correction factor (slope and offset) was applied during the data-processing stage.

The expanded measurement uncertainty (k = 2) of the calibrated sensors was estimated to be approximately ±0.25 °C by combining reference instrument uncertainty, chamber stability, and sensor repeatability. Potential hysteresis effects were assessed by repeating the calibration cycle in reverse order (cool-down), with maximum observed deviation not exceeding 0.1 °C, indicating negligible hysteresis for the intended application.

To evaluate the influence of integration into the glazing system, a post-embedding verification was performed by comparing the calibrated sensor output against the reference thermometer after mounting the PCB inside the glass cavity. An additional offset of 0.1–0.2 °C was observed, attributed to conductive and radiative coupling with the PCB–adhesive–glass stack; this offset is included in the uncertainty budget. No measurable drift was detected during the 24 h experimental period.

3.6. Impact of Miniaturization on Measurement Accuracy

The integration of electronic components inside the glazing cavity imposes strict constraints on device miniaturization, which directly affect thermal accuracy, measurement dynamics, and wireless communication performance. The PCB must remain as small as possible to minimize the thermal disturbance introduced into the cavity. A larger board increases thermal mass and alters the local heat distribution, potentially biasing the temperature readings. As shown in recent compact IoT sensing platforms [27], PCB footprints on the order of 20 × 20 mm are feasible and help reduce thermal inertia. In the present design, the use of ESP8266-based electronics required a slightly larger footprint of 20 × 30 mm, which still fits within the glazing gap while keeping added thermal mass low.

Copper density on the PCB is also a critical factor; high copper coverage increases heat spreading and slows sensor response. To minimize this effect, only essential copper traces were used, and no ground plane was placed directly under the temperature sensor. The selected FR-4 substrate (1.0 mm thickness) (Printed s.r.o., Melnik, Czech Republic) limits thermal conduction while providing sufficient mechanical support. Similarly, SMD package size affects thermal inertia; the BME280 sensor in its LGA package exhibits low thermal mass, enabling a rapid response to changes in air temperature within the glazing cavity.

The adhesive layer used for mechanical fixation introduces additional thermal resistance. To reduce offset effects, a thin, low-conductivity optical adhesive was applied only at the PCB perimeter, avoiding direct contact with the sensor region. Pre- and post-embedding calibration confirmed that adhesive-induced offsets remained within the expected uncertainty range.

From an IoT standpoint, proximity to glass and metallic window frames may detune the PCB antenna and attenuate the Wi-Fi signal. Although the communication distance in the experiment was only 1 m, RSSI remained stable, indicating that the PCB-mounted antenna performed reliably despite the confined environment. Future iterations may employ antenna matching or external antenna routing to further mitigate detuning.

3.7. Influence of Embedded Electronics on the Thermal Field

Embedding active electronics inside the glazing cavity introduces localized heat sources that can affect temperature measurements if they are not properly controlled. Although the ESP8266 and the associated voltage-regulation circuitry dissipate heat during operation, several measures were taken to minimize their impact on the thermal field. First, the boards were operated with a low-duty communication cycle, and during the experiment, the Wi-Fi transmission load was minimal because the access point was located only 1 m from the window chamber. As a result, RF-related heating was negligible, and no observable temperature steps correlated with transmission activity were detected in the recorded data.

To further limit electronics-induced heating, the voltage regulator and power distribution traces were placed away from the sensor region, and the BME280 was positioned on the thermally isolated edge of the PCB. Thermal mass was minimized through reduced copper coverage and a compact 20 × 30 mm board layout. Measurements performed with the system idle (no Wi-Fi transmission) showed no measurable offset compared with fully active operation, indicating that self-heating remained below the sensor's uncertainty range.

The flexible PCB power cable used in the experiment introduced an insignificant thermal bridge due to its very small cross-section, preventing heat conduction from external equipment into the cavity. Overall, the observed thermal gradients were dominated by the glazing behavior rather than the electronics, and the implemented design practices effectively limited hardware-related thermal bias.

4. Results and Discussion

The results from the simulation and the experiment provide valuable insights into the performance of the thermal conductivity in the window, demonstrating the effectiveness of the single-layer glazing design in energy-efficient thermal insulation. The main objective of this study was to validate the simulation model using experimental data, ensuring that the computational predictions were aligned with the observed behavior.

4.1. Results Using CFD Simulation

The CFD simulation revealed a distinct temperature gradient across the glazing chamber, increasing steadily from the cooler interior toward the heated exterior panel. Within the air gap, buoyancy-driven circulation developed as warmer, less dense air rose and cooler, denser air sank, creating a continuous convective loop. Similar natural convection phenomena in enclosed cavities have been widely reported in recent CFD studies, such as [29], confirming that buoyancy effects dominate heat transfer within narrow glazing gaps.

The transient behavior of the simulated sensors further illustrates this effect. Sensor Sim-T1, located closest to the heated panel, recorded the fastest temperature rise, while Sim-T4, positioned near the opposite side, responded more slowly. This delay reflects the time required for heat to propagate through the gas and glass layers. Comparable delays in transient sensor response have been noted in [30], where gas conduction and convection govern thermal dynamics. The outcome of the simulation model can be found in Figure 6.

The graph in Figure 6 illustrates the temperature progression over time as heat was gradually applied to the outer side of the glass panel. The x-axis represents time in seconds, while the y-axis shows temperature in degrees Celsius. The lines labeled Sim-T1, Sim-T2,

Sim-T3, and Sim-T4 correspond to the readings from four simulated sensors placed within the chamber, capturing the heat transfer through the cavity gas from the heated panel to the opposite panel. The black line represents the medium temperature of heating.

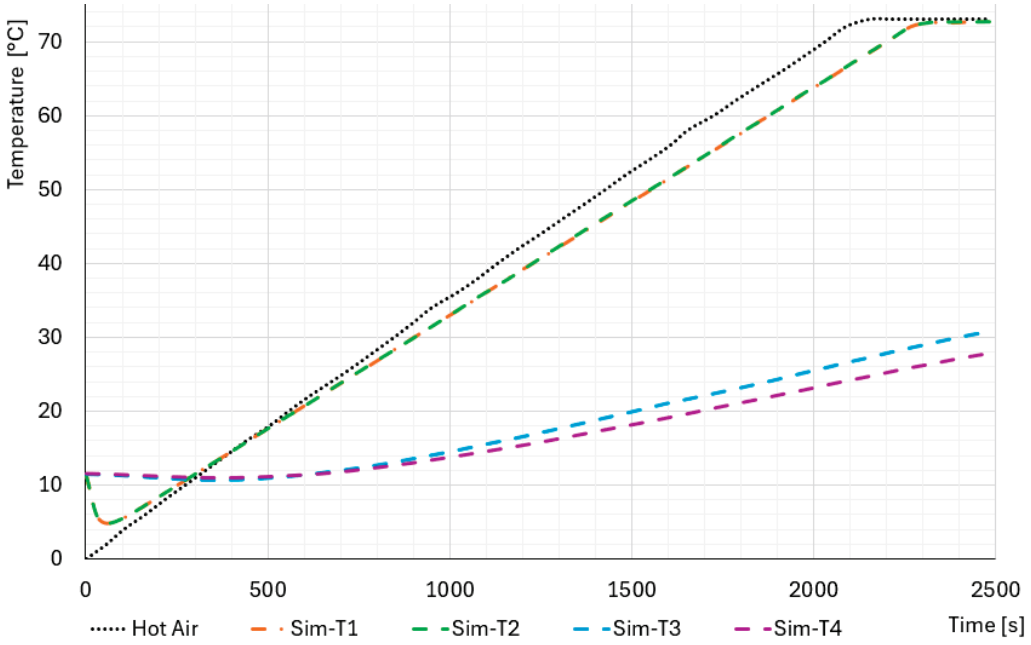


Figure 6. The outcome of the CFD simulation model.

These results not only validate the physical accuracy of the simulation but also align with advanced 3D CFD investigations of window systems, which emphasize the importance of convective and conductive interactions across both glazing cavities and frame structures [31]. Together, the findings reinforce that CFD is a reliable tool for predicting the thermal behavior of glazing systems and for identifying areas—such as cavity convection and frame conduction—that strongly influence insulation performance.

4.2. Results Using Real Experiment

The heating source used in the experiment was a laboratory oven, where one of the side panels was removed and replaced by the tested glazing chamber. In this configuration, the inner side of the chamber was exposed directly to the controlled hot-air environment inside the oven, while the outer side remained in contact with ambient room air, enabling a well-defined temperature gradient across the window. The oven's heating element ensured stable and uniform thermal conditions on the hot side, while no forced convection was applied; thus, heat transfer occurred under natural convection on both sides. The temperature distribution was recorded over time using four sensors placed within the chamber. As heating progressed, the air temperature within the chamber steadily increased, and heat flowed through the glazing as sensed by the temperature sensors. Sensor T1—closest to the heated outer panel—records the most rapid increase, while T4, located near the inner side, responds more slowly. This gradient underscores that ordinary air, with its higher thermal conductivity, transfers heat more quickly across the cavity than inert gases would. In fact, Cho et al. [32] reported that reducing the argon filling rate from 95% to 0% led to roughly a 10.9% decline in thermal performance, corroborating our findings of accelerated heat transfer with air-filled gaps. Additionally, recent work by Rimshin et al. [33] highlights the effectiveness of inert gas fillings and thermal coatings in improving window insulation. These external studies align well with our high-resolution, second-by-second

temperature data, reinforcing that using low-conductivity gases is key for enhancing the thermal efficiency of single-layer glazing systems.

The graph in Figure 7 demonstrates a consistent temperature gradient, with all sensors showing a steady rise in temperature over time. The steeper slopes of the sensor readings compared to experiments with inert gases highlight the reduced insulation properties of air, emphasizing the importance of using low-conductivity gases in glazing systems for enhanced thermal insulation. The data collection process captured detailed thermal performance metrics of the single-layer glazing system, with temperature measurements taken every second for high-resolution results. Calibrated sensors and a data logger ensured accurate and continuous data recording, which was analyzed using software to visualize temperature trends in graphical form.

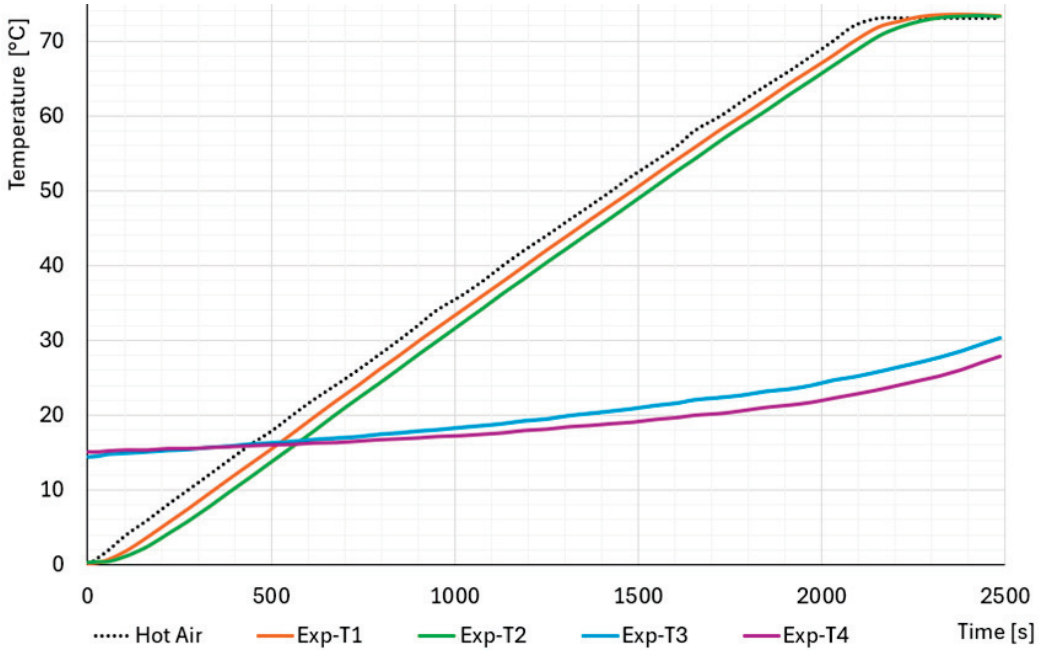


Figure 7. The outcome of the experiment with the heated air.

4.3. Comparison Between CFD Simulation and Real Experiment

The mathematical model and equations used in the simulation were validated through experimental testing to assess their accuracy in predicting thermal behavior. The experiment replicated the simulated conditions by gradually heating one side of the chamber and measuring the heat transfer through the glazing using several temperature sensors. The comparison between simulated and experimental results demonstrated a strong correlation, confirming the reliability of the model. A key finding was the measured temperature difference across the glazing, which plays a crucial role in thermal insulation. This difference directly influences the effectiveness of the window in minimizing heat loss, making it a critical factor in energy-efficient building design.

The simulation results revealed clear temperature gradients across the chamber of the window, with the outer glass panel exposed to the rising temperature, where the outer glass panel reflected the rising temperature, the middle air gap exhibiting moderate temperatures, and the inner glass panel showing the lowest temperatures. These findings were consistently mirrored in the experiment. Temperature measurements taken from the physical prototype confirmed the predicted thermal behavior, with each chamber displaying distinct temperature zones that corresponded closely with the simulation data. The comparison between the simulation and the experiment is shown in Figure 8.

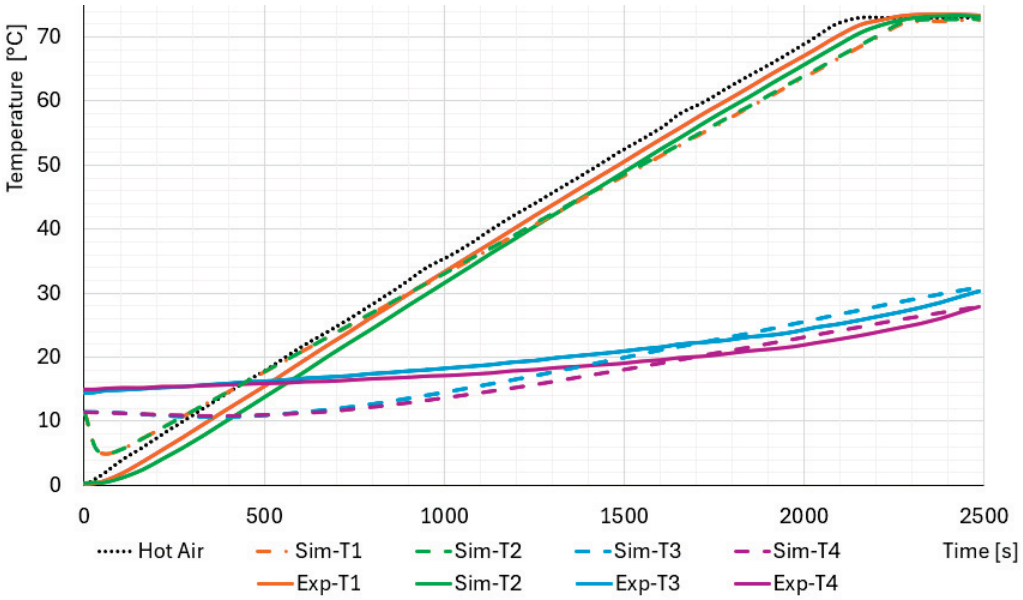


Figure 8. Comparison between the CFD simulation and the experiment.

The graph in Figure 8 presents the temperature evolution over time, illustrating the thermal insulation properties of the window. The x-axis represents time in seconds (0–2500 s), while the y-axis shows temperature in degrees Celsius (0–73 °C). A dotted black line represents the reference temperature of the air, which increases steadily. Experimental (Exp) and simulation (Sim) data are provided for four measurement points (T1–T4), with experimental values shown as solid lines and simulated values as dashed lines. The temperature trends indicate that T1 and T2 experience a rapid increase, closely following the air temperature. In contrast, T3 and T4 exhibit a significantly slower temperature rise, indicating thermal resistance. The close alignment between experimental and simulated data confirms the accuracy of the computational model.

The simulation also revealed that the temperature sensors themselves contributed to heat conduction within the system. The heat from the glass was transferred to the sensors, which then acted as a thermal bridge, conducting heat into the chamber. This unintended behavior introduced an additional pathway for heat transfer, potentially affecting the accuracy of the measurements. Such sensor-induced heat conduction is undesirable, and it can change the thermal dynamics of the glazing system and introduce errors in the experimental data. These findings emphasize the need for careful sensor design and placement to minimize their impact on the system's thermal performance.

4.4. Quantitative Validation Metrics

To provide a rigorous evaluation of the agreement between the CFD simulation and the experimental measurements, a quantitative comparison was performed using common statistical error metrics. For each temperature sensing position (T1–T4), the root-mean-square error (RMSE), mean absolute error (MAE), bias error, and Pearson correlation coefficient r were computed based on paired experimental and simulated temperature data. These metrics allow for the assessment of absolute accuracy (RMSE, MAE), systematic over- or underestimation (bias), and temporal agreement in dynamic response (r).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(T_{sim,i} - T_{exp,i}\right)^2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|T_{sim,i} - T_{exp,i}\right|$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(T_{sim,i} - T_{exp,i}\right)$$

$$r = corr\left(T_{sim}, T_{exp}\right)$$

The results show a high correlation between the simulated and measured temperature curves ($r \geq 0.98$ for all measurement points), indicating that the CFD model accurately reproduces the temporal evolution of the thermal process. The RMSE values range from approximately 2.9 °C to 3.3 °C, and the MAE varies between 2.3 °C and 2.7 °C across all sensors. The bias values remain between −2.1 °C and +1.4 °C depending on the sensor location, demonstrating that the model does not exhibit any substantial systematic offset. Overall, these quantitative results confirm that the simulation provides a reliable approximation of the thermal behavior observed in the experiments.

4.5. Heat Circulation and Thermal Bridges in the Chamber

The simulation results revealed that the spacer frame connecting the two glass panels acts as a significant thermal bridge within the system. Heat conduction through the frame and adjacent glass edges creates an additional pathway for thermal transfer between the inner and outer surfaces. This effect reduces the overall insulation efficiency of the glazing unit, as more heat can escape or enter the chamber. The airflow visualization further illustrates the presence of convection currents that circulate warm air across the chamber, intensifying heat exchange near the edges. These findings highlight the importance of optimizing spacer frame materials and geometry to minimize thermal bridging and improve the thermal performance of glazing systems. The temperature distribution and airflow patterns corresponding to this effect are presented in Figure 9.

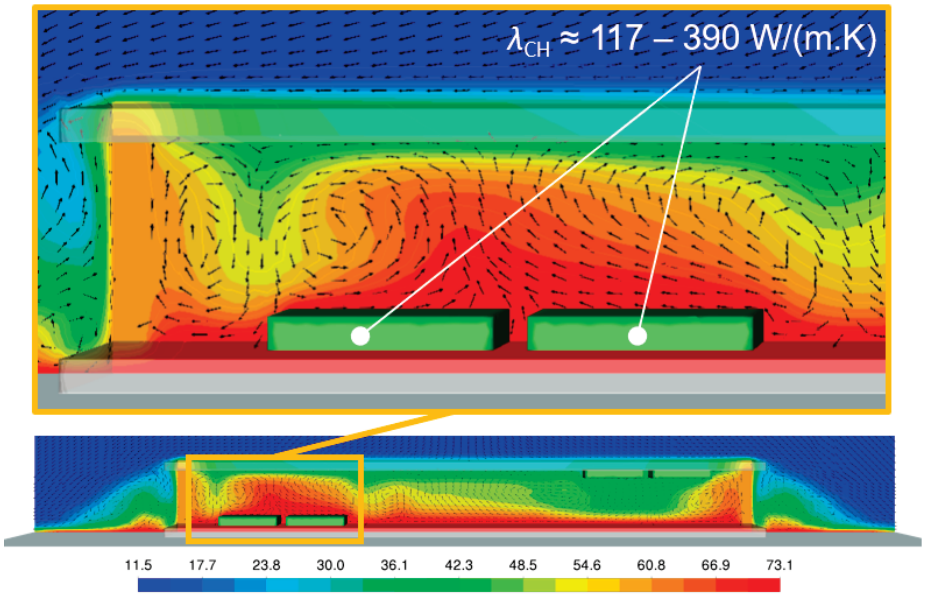


Figure 9. Visualization of the temperature convection of the CFD simulation.

A key source of discrepancy between the simulation and the experiment arises from the representation of the temperature sensor. In the simulation, the sensor is modeled as a full board, whereas in reality, it has a thickness of only 1 mm. This difference introduces potential errors, as the real sensor possesses thermal mass and can conduct heat between the glass and chamber air, effectively acting as a thermal bridge. Its material composition also

enables it to absorb and store heat, slowing its response time and influencing transient readings. Additionally, physical contact with the glass introduces localized thermal resistance not captured in the numerical model. Prior studies confirm that sensor design and thermal inertia significantly affect accuracy; for example, Taler [34] showed that the presence of a thermowell or added mass increases thermal capacitance and delays response, while Małek and Koczyk [35] demonstrated that even small sensor inaccuracies can substantially alter the interpretation of thermal experiments.

Despite certain limitations, the close correlation between simulation and experimental outcomes validates the proposed smart window design. The study revealed areas for refinement, including mitigating the impact of the higher thermal conductivity of air compared to inert gas fills and minimizing the thermal influence of sensors acting as heat bridges. Beyond confirming thermal performance, the integration of IoT functionality enables continuous real-time monitoring, wireless data transmission, and seamless connection with building automation or HVAC systems. This combination of accurate thermal characterization and IoT connectivity demonstrates the potential for scalable deployment in smart homes and buildings, contributing to long-term energy efficiency and sustainable living practices.

5. Conclusions

The integration of multiple sensors into a single IoT-enabled device embedded within glazing represents a significant advancement in building technology and energy management. By combining computational simulations with controlled experiments, this research provides valuable insights into the thermal behavior of glazing materials and the influence of sensor placement on overall performance. The experiments, conducted with air as the heating medium, revealed that the absence of inert gases such as argon or krypton significantly reduces insulation effectiveness due to air's higher thermal conductivity. Moreover, the study highlighted that the sensors themselves acted as thermal bridges, reinforcing the need for improved sensor design to minimize unintended heat transfer.

Despite these challenges, the strong correlation between experimental and simulated results validates the scalable design of the proposed system. Crucially, embedding a temperature-sensing IoT device within glazing extends its function beyond passive insulation toward active, data-driven building management. Real-time monitoring enables continuous performance assessment, cloud-based analytics, and seamless integration with HVAC control systems, ensuring adaptive energy optimization in smart homes and buildings.

This research demonstrates the feasibility of transforming glazing into an intelligent, connected subsystem within the Internet of Things ecosystem. Such integration not only enhances thermal insulation but also contributes to holistic environmental monitoring, supporting energy-efficient, sustainable, and resilient living spaces. The proposed approach lays the foundation for next-generation smart building solutions, where materials and embedded devices collaborate to actively reduce energy consumption and promote a more sustainable built environment. Future work will focus on testing the connectivity, stability, and operational limits of the device within IoT infrastructures, ensuring robust long-term performance in real-world applications.

Author Contributions: V.M.: conceptualization of the study, methodology development, and manuscript writing. J.V.: administration, reviewing and editing. M.A.: system design, integration of multiple sensors, and experimental setup. P.D.: software implementation, data acquisition, and validation of sensor measurements. P.S.: hardware development, sensor placement optimization, and system calibration. S.D.: data processing, statistical analysis, and interpretation of results. L.K.: writing, reviewing and editing, literature research, and discussion of findings. A.M.: administration, resource allocation, reviewing, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Internal Grant Agency of Tomas Bata University in Zlín, under the project No. IGA/CebiaTech/2024/002.

Data Availability Statement: The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge the use of OpenAI's ChatGPT (GPT-5) to support the preparation of this manuscript. The tool was employed for language refinement, improving textual coherence, and standardizing references in accordance with the journal's guidelines. All outputs were thoroughly reviewed, revised, and validated by the authors, who assume full responsibility for the content and integrity of the publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wang, Z.; Wang, Y.; Gao, C.; Wang, F.; Lin, T.; Chen, Y. An adaptive sliding window for anomaly detection of time series in wireless sensor networks. *Wirel. Netw.* **2022**, *28*, 393–411. [CrossRef]
2. Xiang, X.; Li, K.; Huang, B.; Cao, Y. A multi-sensor data-fusion method based on cloud model and improved evidence theory. *Sensors* **2022**, *22*, 5902. [CrossRef]
3. Mansour, A.; Chen, W. SUNS: A user-friendly scheme for seamless and ubiquitous navigation based on an enhanced indoor-outdoor environmental awareness approach. *Remote Sens.* **2022**, *14*, 5263. [CrossRef]
4. Hamad, R.A.; Hidalgo, A.S.; Bouguelia, M.-R.; Estevez, M.E.; Quero, J.M. Efficient activity recognition in smart homes using delayed fuzzy temporal windows on binary sensors. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 387–395. [CrossRef]
5. Zheng, Y.; Si, P.; Zhang, Y.; Shi, L.; Huang, C.; Huang, D.; Jin, Z. Study on the effect of radiant insulation panel in cavity on the thermal performance of broken-bridge aluminum window frame. *Buildings* **2023**, *13*, 58. [CrossRef]
6. Sun, Z.; Cao, Y.; Wang, X.; Yu, J. Multi-objective optimization design for windows and shading configuration: Considering energy consumption, thermal environment, visual performance and sound insulation effect. *Int. J. Energy Environ. Eng.* **2021**, *12*, 805–836. [CrossRef]
7. Kayastha, S.; Upadhyaya, P. Design and implementation of a cost-efficient smart home system with Raspberry Pi and cloud services. In Proceedings of the 2019 Artificial Intelligence for Transforming Business and Society (AITB), Kathmandu, Nepal, 21–23 November 2019; pp. 1–7. [CrossRef]
8. Pham, Q.-Q.; Ta, Q.-B.; Park, J.-H.; Kim, J.-T. Raspberry Pi platform wireless sensor node for low-frequency impedance responses of PZT interface. *Sensors* **2022**, *22*, 9592. [CrossRef] [PubMed]
9. Andreas, C.R.A.; Putra, H.W.; Hanafiah, N.; Surjarwo, S.; Wibisurya, A. Door security system for home monitoring based on ESP32. *Procedia Comput. Sci.* **2019**, *157*, 673–682. [CrossRef]
10. Lyu, Y.; Liu, W.; Su, H.; Wu, X. Numerical analysis on the advantages of evacuated gap insulation of vacuum-water flow window in building energy saving under various climates. *Energy* **2019**, *174*, 353–364. [CrossRef]
11. Baek, K.; Kim, L. Optimum design and energy performance of hybrid triple glazing system with vacuum and carbon dioxide filled gap. *Sustainability* **2019**, *11*, 5543. [CrossRef]
12. Fantucci, S.; Garbaccio, S.; Lorenzati, A.; Perino, M. Thermo-economic analysis of building energy retrofits using VIP—Vacuum insulation panels. *Energy Build.* **2019**, *199*, 269–279. [CrossRef]
13. Thie, C.; Quallen, S.; Ibrahim, A.; Xing, T.; Johnson, B. Study of energy saving using silica aerogel insulation in a residential building. *Gels* **2023**, *9*, 86. [CrossRef] [PubMed]
14. Mehdizadeh-Rad, H.; Choudhry, T.A.; Ng, A.W.M.; Rajabi, Z.; Rais, M.F.; Zia, A.; Tariq, M.A.U.R. An energy performance evaluation of commercially available window glazing in Darwin's tropical climate. *Sustainability* **2022**, *14*, 2394. [CrossRef]
15. Cho, S.; Kim, S.-H. Analysis of the performance of vacuum glazing in office buildings in Korea: Simulation and experimental studies. *Sustainability* **2017**, *9*, 936. [CrossRef]
16. Ali, H.; Hayat, N.; Farukh, F.; Imran, S.; Kamran, M.; Ali, H. Key design features of multi-vacuum glazing for windows: A review. *Therm. Sci.* **2017**, *21*, 2673–2687. [CrossRef]
17. Shi, Y.; Chen, X.; Sun, C.; Xia, X.-L. Temperature-dependent thermal conductivity and absorption coefficient identification of quartz window up to 1100 K. *J. Therm. Sci.* **2023**, *32*, 44–58. [CrossRef]
18. Hodgson, M.; Roy, S.; Roskilly, A.P.; Smallbone, A. The performance and efficiency of novel oxy-hydrogen-argon gas power cycles for zero emission power generation. *Energy Convers. Manag.* **2021**, *244*, 114510. [CrossRef]

19. Nsaif, M.A.; Jali, J.M.; Baccar, M. Experimental investigation of the thermal performance of triple glazed windows integrated with PCM and low-e glass. *Int. J. Heat Technol.* **2024**, *42*, 1735–1743. [CrossRef]
20. Xu, X.; Xie, J.; Zhang, X.; Chen, G.; Liu, J. A new validated TRNSYS module for phase change material-filled multi-glazed windows. *Appl. Therm. Eng.* **2025**, *258*, 124706. [CrossRef]
21. Zhang, C.; Yang, R.; Lu, Y.; Arici, M.; Ma, Y.; Yang, X.; Qi, Z.; Li, D. Parametric research on thermal and optical properties of solid-solid phase change material packaged in glazing windows. *J. Energy Storage* **2024**, *83*, 110562. [CrossRef]
22. Ranđelović, D.; Vorotović, G.; Bengin, A.; Petrović, P. Quadcopter altitude estimation using low-cost barometric, infrared, ultrasonic, and LIDAR sensors. *FME Trans.* **2021**, *49*, 21–28. [CrossRef]
23. Shevchenko, G.V.; Glubokov, N.A.; Yupashevsky, A.V.; Kazmina, A.S. Air flow sensor based on environmental sensor BME280. In Proceedings of the 21st International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM), Erlagol, Russia, 29 June–3 July 2020; pp. 1–5. [CrossRef]
24. Bianchi, C.; Loureiro, J.; Duarte, P.; Marques, J.; Figueira, J.; Ropio, I.; Ferreira, I. V2O5 thin films for flexible and high sensitivity transparent temperature sensor. *Adv. Mater. Technol.* **2016**, *1*, 1600077. [CrossRef]
25. Bornemann, S.; Lang, W. Considerations and limits of embedding sensor nodes for structural health monitoring into fiber metal laminates. *Sensors* **2022**, *22*, 4511. [CrossRef]
26. Zhou, Y. Artificial neural network-based smart aerogel glazing in low-energy buildings: A state-of-the-art review. *iScience* **2021**, *24*, 103420. [CrossRef] [PubMed]
27. Osorio-de-la-Rosa, E.; Valdez-Hernández, M.; Vázquez-Castillo, J.; Franco-de-la-Cruz, A.; Woo-García, R.; Castillo-Atoche, A.; La-Rosa, R. Plant microbial fuel cells as a bioenergy source used in precision beekeeping. *Sustain. Energy Technol. Assess.* **2023**, *60*, 103499. [CrossRef]
28. *ISO/IEC 17025*; General Requirements for the Competence of Testing and Calibration Laboratories. ISO/IEC: Geneva, Switzerland, 2017.
29. Zhang, Y.; Zhang, M.; Xiong, J.; Mao, G.; Qi, Y. Computational fluid dynamics for cavity natural heat convection: Numerical analysis and optimization in greenhouse application. *Adv. Math. Phys.* **2023**, *2023*, 1–11. [CrossRef]
30. Samuel, Y.; Confrey, T.; Callaghan, D.; Kent, N.; Nolan, C. CFD analysis of thermal and flow physics in buildings utilizing smart glazing for mitigation of solar gain. In Proceedings of the 5th Thermal and Fluids Engineering Conference (TFEC), New Orleans, LA, USA, 5–8 April 2020; Begell House: Danbury, CT, USA, 2020; pp. 151–161. [CrossRef]
31. Basok, B.; Novikov, V.; Pavlenko, A.; Davydenko, B.; Koshlak, H.; Goncharuk, S.; Lysenko, O. CFD simulation of heat transfer through a window frame. *Rep. Struct. Eng.* **2024**, *26*, 56–64. [CrossRef]
32. Cho, K.; Cho, D.; Koo, B.; Yun, Y. Thermal performance analysis of windows, based on argon gas percentages between window glasses. *Buildings* **2023**, *13*, 2935. [CrossRef]
33. Rimshin, V.; Khamrakulov, R.; Alikabulov, S.; Radjabov, Y.; Abdurakhmonov, A.; Mirazimova, G.; Jamolova, M. Study on the possibilities of increasing the effectiveness of thermal insulation of enclosing structures in window openings using low-emission coatings and films. *E3S Web Conf.* **2024**, *563*, 02024. [CrossRef]
34. Taler, D.; Sobota, T.; Jaremkiewicz, M.; Taler, J. Influence of the thermometer inertia on the quality of temperature control in a hot liquid tank heated with electric energy. *Energies* **2020**, *13*, 4039. [CrossRef]
35. Małek, M.T.; Koczyk, H. Influence of temperature sensor (Pt100) accuracy on the interpretation of experimental results of measuring temperature on the surface. *Civ. Environ. Eng. Rep.* **2024**, *34*, 1–21. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

Intelligent Control Approaches for Warehouse Performance Optimisation in Industry 4.0 Using Machine Learning

Ádám Francuz and Tamás Bányai *

Institute of Logistics, University of Miskolc, 3515 Miskolc, Hungary; francuzadam13@gmail.com
* Correspondence: tamas.banyai@uni-miskolc.hu

Abstract

In conventional logistics optimization problems, an objective function describes the relationship between parameters. However, in many industrial practices, such a relationship is unknown, and only observational data is available. The objective of the research is to use machine learning-based regression models to uncover patterns in the warehousing dataset and use them to generate an accurate objective function. The models are not only suitable for prediction, but also for interpreting the effect of input variables. This data-driven approach is consistent with the automated, intelligent systems of Industry 4.0, while Industry 5.0 provides opportunities for sustainable, flexible, and collaborative development. In this research, machine learning (ML) models were tested on a fictional dataset using Automated Machine Learning (AutoML), through which Light Gradient Boosting Machine (LightGBM) was selected as the best method (R^2 = 0.994). Feature Importance and Partial Dependence Plots revealed the key factors influencing storage performance and their functional relationships. Defining performance as a cost indicator allowed us to interpret optimization as cost minimization, demonstrating that ML-based methods can uncover hidden patterns and support efficiency improvements in warehousing. The proposed approach not only achieves outstanding predictive accuracy, but also transforms model outputs into actionable, interpretable insights for warehouse optimization. By combining automation, interpretability, and optimization, this research advances the practical realization of intelligent warehouse systems in the era of Industry 4.0.

Keywords: warehouse performance; optimization; regression; machine learning; feature importance

1. Introduction

Industry 4.0 (I4.0) technologies are increasingly being deployed across diverse sectors, ranging from fire and safety management [1] to traffic control systems, where real-time data and automation enhance decision-making and operational reliability. These applications highlight how intelligent systems and advanced analytics can transform traditional processes into adaptive, data-driven solutions. Within logistics, the integration of Industry 4.0 is particularly relevant, as warehousing represents a critical link between production and distribution. Modern warehouses face complex challenges such as fluctuating demand, dynamic material flows, and rising efficiency requirements.

By leveraging ML, IoT, and automation, I4.0 opens new opportunities for optimizing storage performance and resource allocation. Consequently, warehouse management becomes a prime field for exploring how data-driven methods can redefine cost-efficiency and adaptability in industrial practice.

Within this frame, we review selected articles to highlight the importance of the research field and identify research gaps. The review focuses on peer-reviewed papers published between 2000 and 2025, identified in Scopus and Google Scholar using the keywords "warehouse optimization," "machine learning in logistics," and "Industry 4.0 warehouse performance." Priority was given to studies that applied quantitative optimization or AI-based approaches in warehousing contexts. This addition strengthens the methodological transparency of the literature review.

The rapid expansion of e-commerce has fundamentally reshaped logistics and warehousing systems, compelling manufacturing companies to shorten the order-to-delivery cycle and adopt intelligent, data-driven optimization strategies [2]. Reinforcement learning (RL) and soft computing methods have been successfully applied to warehouse automation, supporting dynamic scheduling and coordination of Automated Guided Vehicles (AGVs) and stackers [2]. In parallel, the Industrial Internet of Things (IIoT) and Edge Intelligence (EI) paradigms have enabled real-time, distributed decision-making closer to the source of data, enhancing reliability and responsiveness compared to cloud-centric architectures [3,4]. These developments illustrate the growing importance of IIoT intelligence for digital manufacturing transformation and large-scale industrial integration [4]. Moreover, integrating RL with time-series forecasting and deep learning techniques has proven effective in improving demand prediction and inventory management, enabling adaptive responses to market volatility [5]. Building on these advances, the present study applies machine learning-based regression models and automated modeling techniques to uncover hidden performance patterns in warehouse operations, offering interpretable, cost-oriented optimization insights consistent with the principles of Industry 4.0 and Industry 5.0.

Cost reduction and efficiency improvement are strategic priorities for manufacturing companies operating in competitive markets. While certain expenditures are inherently tied to value-creating activities, such as production and quality assurance, and thus considered essential, there are numerous sub-processes within manufacturing and supply chain (SC) operations that offer potential for reorganization and optimization. By identifying and refining these areas, companies can achieve measurable cost savings. Even a reduction of a few tenths of a percentage point in operational costs can translate into substantial financial benefits, underscoring the importance of exploiting optimization opportunities through advanced analytical and algorithmic approaches.

The efficient organization of finished goods warehouses plays a critical role in overall SC performance. Numerous configurations exist for structuring SCs and managing the flow of finished products, typically involving the transfer of goods from manufacturing plants to distribution centers, followed by delivery to retailers. This study focuses on the performance optimization of finished goods warehouses by analyzing a dataset containing key performance indicators from various manufacturing sites. The dataset includes multiple descriptive parameters characterizing warehouse operations. The primary objective variable in the optimization process is the quantity of outgoing materials, which serves as a proxy for warehouse throughput and operational efficiency.

Warehouse performance optimisation can be approached via several parameters and broken down into several sub-processes. According to one study, the costliest process in warehouse operations is order picking, which can account for up to 55% of total operating costs [6]. The general approach is to determine warehouse performance based on the distance travelled by various material handling equipment. This can be optimally achieved by designing a layout tailored to the specific warehouse [7]. The minimum distance travelled can also be achieved with different storage strategies, as the picking route can be reduced by dynamically changing the destination. The random assignment method results in high space utilisation (or low space requirements), but this is accompanied by

an increase in distances. This operation can only be implemented with IT support. If the storage location can be freely selected, a layout can be created in which the utilisation rate varies significantly in different areas of the warehouse [8]. According to one study, the two strategies are almost equally efficient [9]. The Cube Per Order Index (COI) strategy was developed to enable efficient storage. Its aim is to place products in the warehouse based on their picking frequency and space requirements, so that fast-moving products are closer to the picking area, thus saving travel time. However, the method does not consider multi-command orders, i.e., product–product relationships that arise during an order, which can significantly impair the performance of the method [10]. The Order Oriented Slotting (OOS) method was developed to eliminate these errors. It focuses on which products occur frequently in an order and arranges them in nearby storage locations, thus reducing the distance travelled during order picking [11].

In addition to layout planning methods, other approaches have also been developed: the batching method examines how picking processes can be coordinated for small orders, thereby reducing travel time [12]. Using the zoning method, the warehouse is divided into different areas, and SKUs belonging to the same product group are stored close to each other. Compared to the batching method, this procedure does not have a significant impact on the performance of warehousing systems [13].

Many procedures and methodologies have been developed to improve warehouse performance efficiency. However, with the spread of artificial intelligence, new and significantly more efficient methodologies have been created that offer more comprehensive solutions that are better adapted to individual needs than general and theoretical procedures, utilising technological advances. Due to growing customer demands and constantly evolving global competition, one of the most important indicators is the choice of warehousing strategy, which is why the dynamic storage location assignment problem (DSLAP) was defined. Deep Reinforcement Learning (DRL) is used to solve this problem, which helps to find the optimal storage locations to reduce costs within the warehouse. When extending this to a real industrial problem, it was found that after two months, warehouse costs were reduced by 6.3% [14], which is a significant saving, given that the method did not require any significant investment.

In the field of SCM, the use of artificial intelligence is most prevalent in demand forecasting. Based on recent years, numerous ML [15], deep learning [16] and time series [17] models have been examined for the implementation of forecasting. It has been found that AI-based forecasting is increasingly replacing traditional statistical models, and companies are investing more and more in AI technologies to optimise their entire SC [18]. Emerging research covers various industries, including the automotive industry [19], healthcare [20] and retail [21].

Several methodologies have been developed to implement order batching. A genetic algorithm-based metaheuristic approach has been developed to minimize picking distances, which is both fast and efficient for large warehouses [22]. Genetic algorithms can be supplemented with weighted association rules, so order batching and thus the avoidance of delays can also be implemented using a hybrid model [23].

Various formulations have been developed directly for route optimization, in which the layout and other factors are considered constant, so implementation must be carried out with the available resources. The preferred solution technique is the “S-shape” heuristic, as it can be used frequently due to its simplicity [24], while the Largest Gap, Return, and Composite methods have also become popular [25]. In addition to route determination, optimization can also be implemented in other ways, as the results of strategic-level predictive planning based on storage technologies, material handling systems, and picking strategies can approach the efficiency of general optimization [26].

In warehousing processes, many sub-processes can be replaced by robotics for automation purposes. The use of robotics is most often complemented by some form of AI-based decision-making, for which Reinforcement Learning-based algorithms [27] or Deep Learning-based methods [28] may be suitable choices for real-time control of robots to achieve maximum performance.

Based on the above literature review, the spread of artificial intelligence in the field of SCM is widespread and its popularity is constantly growing. One element of the SC is warehousing, for which numerous methodologies and algorithms have also been developed, and case studies confirm the importance of this research topic. The review of the literature revealed no studies employing regression models for pattern recognition rather than prediction in the context of SCM; consequently, the research objective can be regarded as both valid and significant for future applications.

The aim of the research is to employ a less common method rather than general and well-known optimization techniques for maximizing the material flow of warehouse performance. The primary goal of ML-based regression methods is prediction, i.e., the determination of unknown states. In such models, the recognition of relationships between variables is essential, as these parameters influence the value of the target variable. These relationships could be expressed and subsequently used to determine the objective function.

The novelty of the approach lies in the application of automated ML techniques for selecting the most suitable regression model and in the use of interpretable methods, such as Feature Importance and Partial Dependence Plots, to reveal hidden patterns in warehousing processes. Furthermore, it is demonstrated how these insights can be translated into mathematical functions defining cost-based performance, thereby offering a new perspective for optimization in industrial practice.

This paper fits within the scope of *Future Internet* as it demonstrates how ML and automated modeling can transform warehouse operations into data-driven, networked, and intelligent systems. By linking logistics optimization with Industry 4.0 and Industry 5.0 concepts, the research contributes to the broader vision of Future Internet, where industrial processes are increasingly interconnected, adaptive, and sustainable. Thus, the study aligns with the journal's focus on advancing digital and internet-based technologies across different domains.

The remainder of this paper is structured as follows. Section 2 describes the research methodology, including the applied optimization and ML techniques, as well as the dataset and modeling procedures. Section 3 presents the results obtained from the AutoML modeling, feature interpretation, and optimization modeling. Section 4 discusses the implications, limitations, and potential industrial relevance of the findings. Finally, Section 5 concludes the study and outlines directions for future research. Figure 1 shows the graphical abstract of the article. The figure illustrates the main steps of the proposed methodology for warehouse performance optimization: dataset preparation and preprocessing, AutoML model selection (LightGBM), feature interpretation using Feature Importance and SHAP/PDP analysis, and formulation of cost-based mathematical optimization functions. The workflow demonstrates how data-driven modeling supports interpretable and efficient decision-making in Industry 4.0 warehousing.

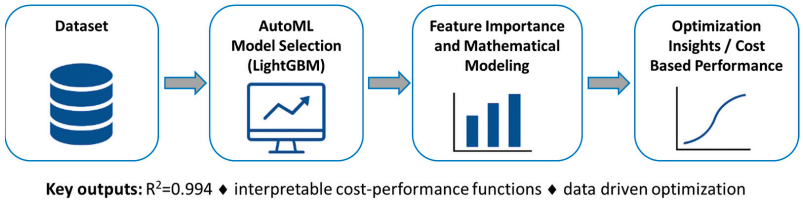


Figure 1. Graphical abstract of the research framework (source: own elaboration).

By integrating automated modeling and interpretable optimization, this research advances the development of smart, transparent, and cost-efficient warehouse systems aligned with the principles of Industry 4.0.

2. Materials and Methods

This section presents the modeling steps of the research. Section 2.1 provides an overview of optimization methods in logistics, while Section 2.2 reviews the main ML models applied in this context, thereby placing this work into a broader methodological framework. Section 2.3 presents the dataset used for the research and its general properties, focusing on the target variable and the most important parameters. Section 2.4 discusses the correlations between the indicators, their meaning and possible consequences, and their potential uses. For effective modeling, the AutoML method is used, the theory and results of which are explained in Section 2.5. Finally, the model with the best evaluation is presented in Section 2.6.

2.1. Optimization Methods in Logistics

The subject of logistics process optimisation is broad and offers various methodologies for solving numerous problems to achieve more efficient, faster and more cost-effective operations. According to the general approach, various logistics problems can be understood as mathematical optimisation, the simplest case of which is linear programming, where both the objective function and the constraints are specified in linear form. A more advanced solution to mathematical optimisation is dynamic programming, which is a recursive approach, meaning that a problem consists of several decision steps that build on each other.

The complexity of a logistics problem often exceeds the capabilities of mathematical modelling. Heuristic algorithms were developed to solve this problem, as they can provide optimal solutions to more complex, non-linear and multi-variable problems. Metaheuristic frameworks can be applied to various types of optimisation problems, the best-known types being genetic algorithms (GA), tabu search, simulated annealing and Ant Colony Optimisation (ACO) [29].

Simulation modelling in logistics optimisation allows the operation of complex systems to be tracked and analysed without having to implement various activities. These procedures are particularly useful as they help to predict different scenarios and the potential risks of events, while risk analysis can be used to determine the optimal operating conditions. The most widely known simulation methods are discrete event simulation (DES), which is a time-based method that follows the sequence of events [30], and Monte Carlo simulation, which is a statistical method that uses random inputs and models using thousands or millions of different runs [31].

In addition to programming-based optimisation procedures, organisational and logic-based process development procedures have also been created. The Lean and Six Sigma methodologies have identified problems that can be observed during operation and addressed with simple solutions, including loss reduction, quality improvement and standardisation [32]. Parallel to the development of digital technologies, modern methodologies requiring IT support have been developed to facilitate optimisation, resulting in the emergence of Digital Twin technology [33] and IoT integration [34].

Nowadays, the popularity of artificial intelligence and ML-based techniques remains robust, and their use is becoming increasingly widespread in many industries. The application of predictive techniques can be particularly important in logistics processes; for example, demand forecasting can significantly support warehousing and capacity planning tasks, as it can prevent excessively high inventory levels or product shortages [35].

Various object detection solutions can be used to improve warehousing processes, helping to identify specific products and analyse capacities [36]. In addition to warehousing, these solutions can also support the optimisation of other processes, such as Reinforcement Learning methods, which are increasingly used for route planning and optimizing transportation networks [37]. In this research, ML-based regression techniques were used, but not as the primary goal of use, rather to support an alternative approach.

Effectiveness and novelty comparing of prior approaches to warehouse optimisation and this research are shown in Table 1.

Table 1. Effectiveness and novelty: comparison of prior approaches to warehouse optimisation and this research.

Approach	Main Focus	Limitations	Novelty
Layout design & distance minimisation [14–17]	Reduce travel distances of material handling equipment	Static, layout-specific, limited adaptability	Goes beyond static layouts by extracting dynamic patterns from data
Slotting strategies (COI, OOS) [18,19]	Product placement based on frequency or co-occurrence	Limited to order relations, ignores other factors	Considers multiple parameters simultaneously via regression
Batching & zoning [20,21]	Coordinating small orders, grouping SKUs	Limited overall impact on performance	Broader optimisation perspective using data-driven models
Metaheuristics (GA, hybrid models) [30,31]	Optimising picking and batching routes	High computational cost, less interpretable	Our approach emphasizes interpretability and cost functions
Routing heuristics (S-shape, Largest Gap, etc.) [32,33]	Picker route optimisation	Simplistic, not globally optimal	Proposes mathematical functions derived from ML for optimisation
Robotics + AI (RL, DL) [35,36]	Automation and real-time control	Requires heavy investment, complex integration	Provides a lightweight, interpretable alternative for optimisation
This research	Pattern recognition via regression models (AutoML + LightGBM)	Needs validation on real data, causality not guaranteed	Transforms regression from prediction to pattern extraction, formulating an optimisation-oriented, interpretable objective function

2.2. Machine Learning Models

The purpose of ML-based models is to recognise patterns and correlations in the data sets under examination, which are then used to make predictions, decisions or classifications. Based on the information collected, they can apply the correlations to new, unknown cases and make estimates for these situations. ML models can be divided into two main groups: supervised learning and unsupervised learning.

The most important types of supervised ML are regression models. In this type of model, the predicted target variable is a numerical value, meaning that the model estimates a specific numerical value based on the specified independent parameters. The independent variables can be either numerical or categorical. In the case of categorical variables, they must be transformed into numerical values using encoding methods. In this research, the target variable depends on several variable values, so initially it is possible to write the predictive model as a multivariate linear regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \quad (1)$$

where y is the dependent value, $x_1, x_2, \ldots, x_n$ are the independent values, β_0 is the intercept (constant term), $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients, and ε is the error term or residual.

There are numerous statistical indicators for evaluating regression models, which examine from different perspectives the quality with which a model can predict unknown target variables based on the trained data set:

- MAE: Mean Absolute Error
- MSE: Mean Squared Error
- R^2: Coefficient of Determination

In this research, the regression model is not used for prediction but rather for the extraction of patterns identified during model creation, thereby allowing various correlations between the target variable and other parameters to be identified. To extract patterns, it is essential that the evaluation of the model building is positive, i.e., the metrics presented above show sufficiently good ratios in relation to the predictive power of the model. If this evaluation is successful, hidden patterns can be extracted from the model (see Figure 2).

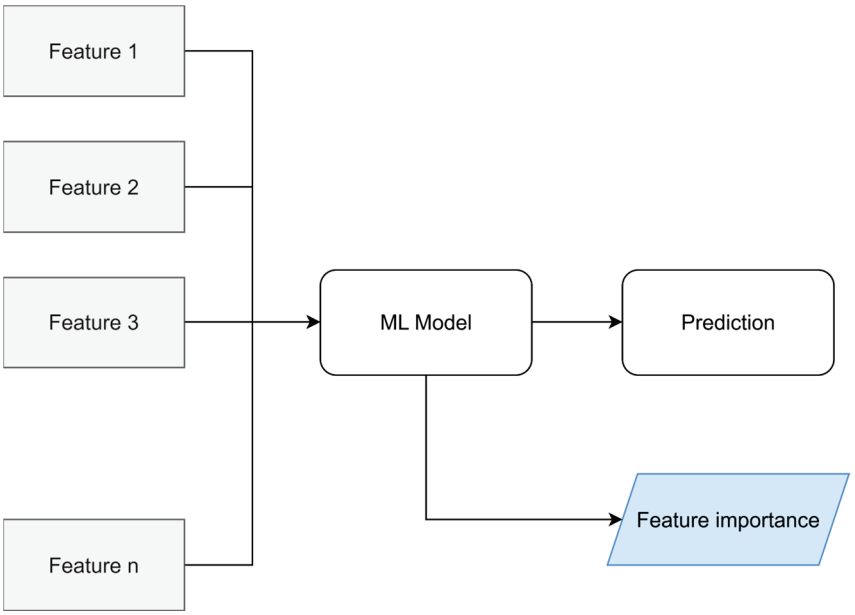


Figure 2. Feature importance extraction from ML models (source: own elaboration).

There are several interpretation techniques in ML that help understand the extent to which different input variables contribute to the value of the target variable. The most widely known technique is Feature Importance, which examines, specifies and ranks the importance of variables in a weighted list.

The SHAP (Shapley Addictive exPlanations) method iterates through all predictions and shows how each variable influences a given prediction. Like the feature importance method, it displays the independent variables in a ranking. The Partial Dependence Plot (PDP) shows how the model responds to the prediction when a categorical parameter is changed. By averaging these evaluations, the average level of change in the target variable resulting from a change in a parameter can be obtained.

These techniques are used in research to determine which parameters most influence the target variable, and the results are expressed in a target function that also includes the degree of influence.

In summary, although several optimization methods have been developed for warehouse operations, there is still a need for approaches that go beyond pure prediction and enable the extraction of hidden patterns from operational data. This perspective opens

up opportunities for constructing optimization-oriented objective functions that are more closely aligned with real industrial processes. Therefore, the main objective of this research is to explore the applicability of ML-based regression models for identifying hidden patterns in warehouse performance data and to formulate an optimization-oriented objective function from the extracted relationships. Beyond demonstrating the predictive capabilities of the models, the aim is to highlight their interpretative power in supporting data-driven decision-making in warehouse management.

2.3. Dataset Introduction

In this research, a dataset containing 25,000 warehouse cases after data cleaning and preprocessing was used. The dataset presents the given warehouse with several characteristics, including both numerical and categorical variables. Initially, no information is available about the relationships between the variables. The dataset is freely available for download and use, ensuring transparency and reproducibility of the research [38].

A total of 20 variables can be identified that may influence the target variable. These include variables describing location (e.g., type of location, zone, regional zone, etc.), parameters characterizing various events (e.g., shipping errors, warehousing errors during a given period), physical characteristics (e.g., geographical, electronic, etc.), and characteristics necessary for warehouse operation (number of employees, number of retailers served, distances, etc.) (Figure 3).

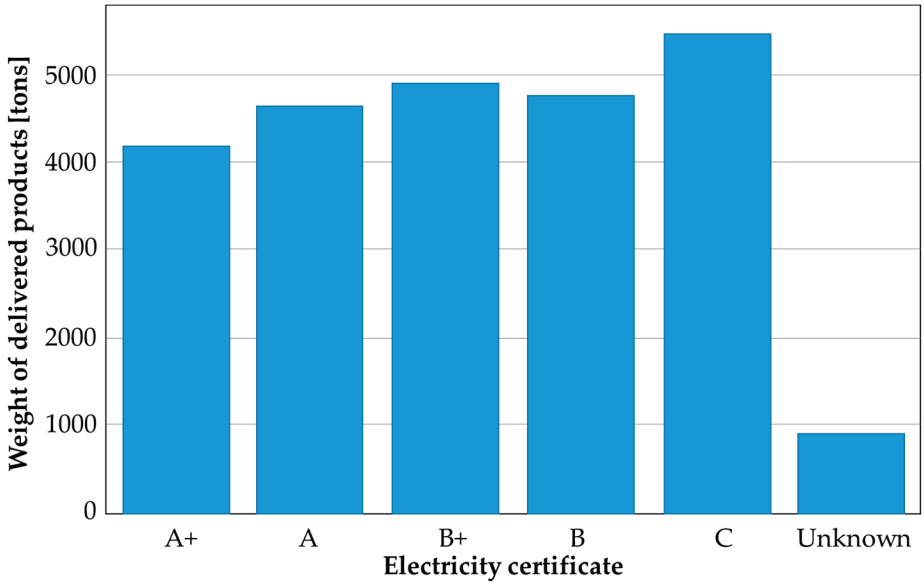


Figure 3. Type of electricity certificate could be important for modelling (source: own elaboration).

The target variable of the model is the weight (in tons) of the products delivered by the warehouse. This can be considered the most relevant parameter, as the goal in this research is to model this parameter with the highest possible quality, and then, using this model and extracting the hidden patterns, determine a possible, optimizable target function.

The main function of the violin plot is to display the information that can be shown by both the box plot and the histogram (Figure 4). The boxplot shows the general statistical characteristics (median, quartiles, and outliers), while the histogram shows the distribution of values. The distribution of the dependent variable values is suitable for building a ML model.

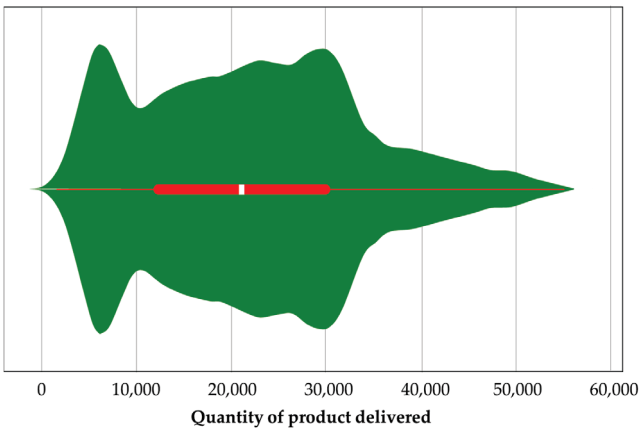


Figure 4. Violinplot of dependent variable (source: own elaboration).

2.4. Exploratory Data Analysis

When recognizing data sets, it is crucial to identify the relationships between variables. Correlation analysis is a suitable choice for this purpose, as it is a statistical indicator that describes the linear relationship between variables on a scale of $[-1, 1]$. This is a good starting point, but the relationship between variables is not necessarily linear, and the analysis can only compare numerical variables; other methods can be used to analyze categorical variables.

Figure 5 shows the correlations between the different variables. The focus is placed on the target variable, the "product_wg_ton" parameter. The target variable has the highest correlation value with another variable. The "storage_issue_reported_l3m" variable shows how many storage errors the given warehouse has reported in the last 3 months. The correlation is close to the maximum, which indicates a very high linear relationship. This relationship is also clearly visible in Figure 6. However, it is important to note that this indicator is not an influencing factor, i.e., increasing errors does not lead to an increase in warehouse performance, but is a consequence, as higher warehouse performance is associated with more operational errors. This observation also applies to the "wh_breakdown_l3m" parameter.

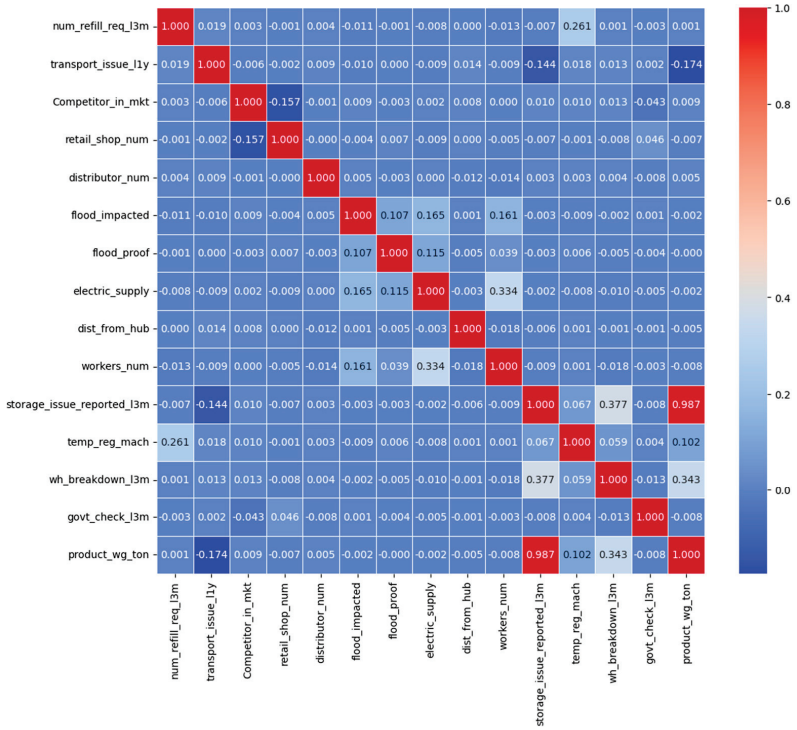


Figure 5. Correlation Matrix (source: own elaboration).

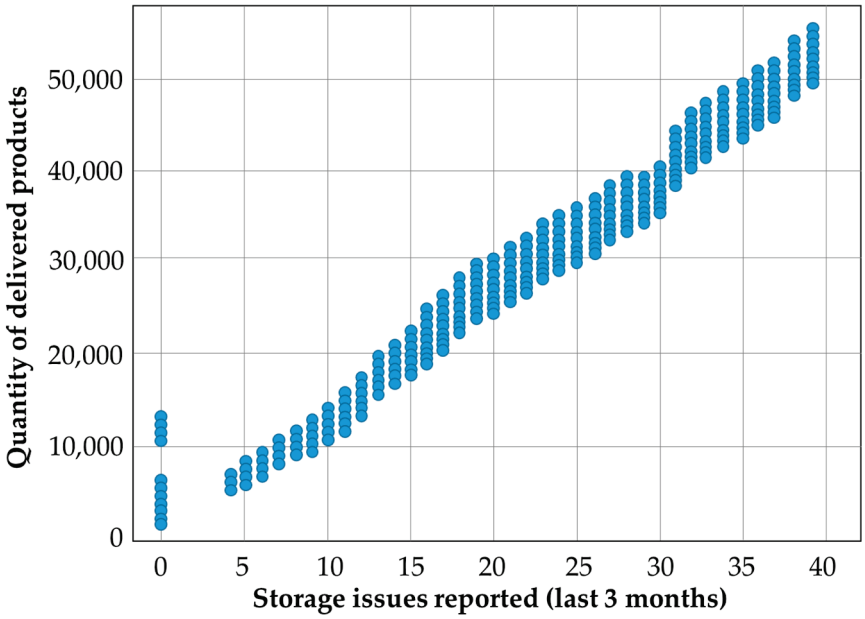


Figure 6. Correlation between dependent variable and storage issues (source: own elaboration).

The goal of this research is to identify the parameters that influence and affect the target variable. This characteristic cannot be determined from correlation indicators; the determination and interpretation of relevant factors require the application of domain-specific knowledge in logistics.

2.5. Applying of Automated Machine Learning

Preparing and training machine learning models is very time-consuming, and each model must be trained separately. Automated Machine Learning (AutoML) began to spread in the 2010s. It is a concept that automates ML processes, thereby reducing the time required and the amount of human intervention in building systems [39]:

- The process starts with data preparation, which includes data collection, data cleaning, and data augmentation.
- Next comes feature engineering, where raw data is transformed into useful features. This involves feature selection, feature extraction, and feature construction. The outcome is the set of features used by the models.
- In model generation, the search space is defined: it can involve traditional models (e.g., SVM, KNN) or deep neural networks (e.g., CNN, RNN). Performance is improved through optimization methods such as hyperparameter optimization and architecture optimization. The process can also include Neural Architecture Search (NAS).
- Finally, in model estimation, efficient evaluation methods are applied to save computational resources: low-fidelity estimation, early stopping, surrogate models, and weight-sharing.

AutoML may be a suitable choice, as the goal is to select the best method for subsequent optimization. After evaluating the models, there are numerous opportunities for further development to improve model building, but the results offered by the AutoML methodology are already a suitable choice for filtering out correlations. The PyCaret open-source library is used for the application, as despite its low-code nature, it allows the automation of the ML processes required for this research.

The authors have chosen PyCaret because it provides a unified and low-code framework for rapid prototyping, model comparison, and reproducible experimentation. While more specialized libraries may exist, PyCaret integrates multiple algorithms and prepro-

cessing steps under a single, standardized. This not only accelerates the experimental cycle but also minimizes the implementation overhead, allowing researchers to focus on the scientific problem rather than technical details. Furthermore, prior familiarity with PyCaret ensured an efficient workflow, reducing the cognitive and time costs associated with learning a new framework. In this context, the choice of PyCaret is scientifically justified by the trade-off between ease of use, reproducibility, and research efficiency.

Based on Table 2, it can be observed during the evaluation that an AutoML model examines several ML models and provides feedback based on various metrics. In the case of models built on the data set, the Light Gradient Boosting Machine (LightGBM) model provides the best evaluation according to all indicators, so this model will be used for this research.

Table 2. Evaluation of AutoML model (source: Python 3.12.11).

	Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE
lightgbm	Light Gradient Boosting Machine	678.8557	839,779.1779	916.0946	0.9938	0.0801	0.0438
gbr	Gradient Boosting Regressor	695.8355	869,759.6419	932.2677	0.9935	0.0837	0.0456
rf	Random Forest Regressor	712.245	934,943.3382	966.6337	0.9931	0.0834	0.0459
xgboost	Extreme Gradient Boosting	713.7591	937,255.2687	967.598	0.993	0.0876	0.0469
et	Extra Trees Regressor	730.8405	1,013,394.922	1006.1808	0.9925	0.0859	0.0469
ridge	Ridge Regression	884.6469	1,309,630.251	1144.0836	0.9903	0.1092	0.0611
llar	Lasso Least Angle Regression	883.8248	1,309,304.182	1143.9377	0.9903	0.1087	0.0609
br	Bayesian Ridge	884.6194	1,309,632.473	1144.0844	0.9903	0.1091	0.0611
lasso	Lasso Regression	884.0397	1,309,316.367	1143.9438	0.9903	0.1089	0.061
lr	Linear Regression	884.7628	1,309,630.904	1144.0847	0.9903	0.1093	0.0612
dt	Decision Tree Regressor	882.6388	1,773,604.005	1331.0611	0.9868	0.1127	0.056
en	Elastic Net	1152.8724	2,577,668.8	1604.7077	0.9809	0.2473	0.0773
ada	AdaBoost Regressor	1413.0806	3,132,509.587	1769.5949	0.9768	0.1617	0.1072
huber	Huber Regressor	1264.9622	3,142,596.727	1767.4949	0.9767	0.4595	0.0853
omp	Orthogonal Matching Pursuit	1353.6249	3,421,633.209	1849.2622	0.9746	0.3484	0.0895
par	Passive Aggressive Regressor	1587.6008	4,461,971.235	2078.8567	0.9668	0.3997	0.1086
knn	K Neighbors Regressor	6103.7732	60,215,414.01	7759.333	0.5534	0.4724	0.459
dummy	Dummy Regressor	9576.893	134,923,670.3	11,615.2979	−0.0008	0.6743	0.7883
lar	Least Angle Regression	91,640.5912	1.116×10^{15}	133,455.673	−860.208	0.8561	6.2092

2.6. LightGBM Model

Gradient Boosting Machine is an ensemble learning technique that typically trains decision trees iteratively, with each new model correcting the errors of the previous ones. The algorithm optimizes predictions by moving in the direction of the gradient of the loss function [40].

In recent years, several improved versions of the GBM algorithm have appeared, the most popular being CatBoost, XGBoost, and LightGBM. The latter model was published by Microsoft Research Asia in 2017. Its goal is to be faster and more efficient than previous GBM implementations, especially with large data sets and high dimensions. Figure 7 shows that the traditional GBM algorithm builds the decision tree level by level, expanding all leaves at each level before moving on to the next level. In contrast, the LightGBM algorithm uses leaf-level growth, meaning that it always expands the leaf that results in the greatest

error reduction [41]. Furthermore, decision trees are also capable of predicting numerical values, so these models can be used to solve regression problems [42].

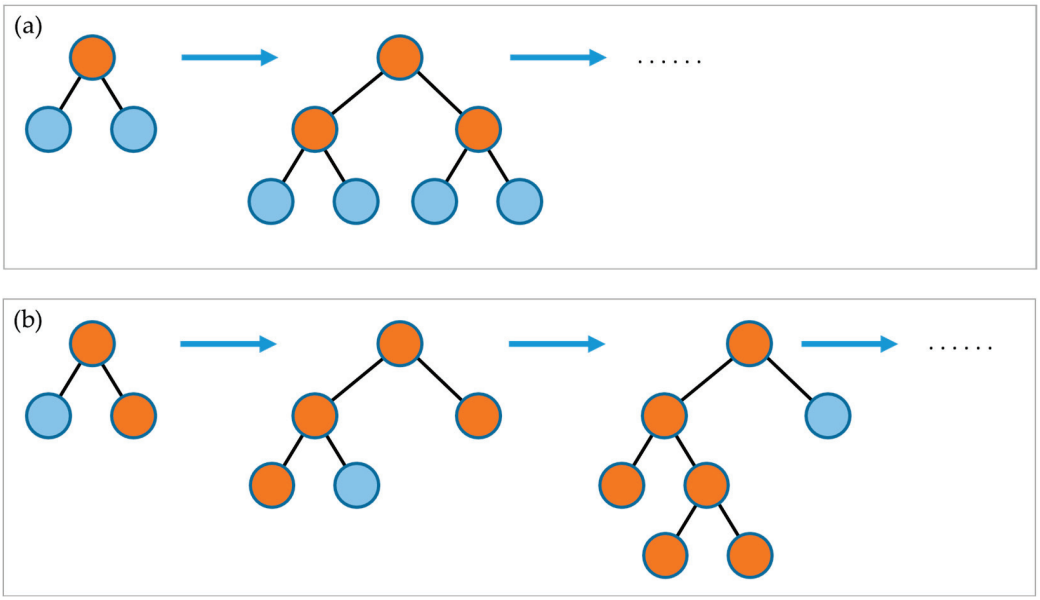


Figure 7. Comparison of level-wise tree growth (**a**) and leaf-wise tree growth (**b**).

For later use, the model was generated using the LightGBM library, where an exceptionally high R^2 value of 0.994 was also obtained, corresponding to the result shown in Table 1.

3. Results

The following chapter presents the main question examined by the research, namely the potential uses of ML methods for logistics performance optimisation. Section 3.1 examines the results of the feature importance and SHAP methods, highlighting the parameters found to be most important. These indicators are examined in Section 3.2, which uses the PDP methodology to examine the consequences of changing a given variable. The resulting methodology is presented in Section 3.3 through mathematical modelling.

3.1. Feature Importance

The objective of this research is for hidden relationships to be extracted from the constructed model and used to develop an optimisation model. To achieve this, a number of methods have been developed and are incorporated in order to explore the widest possible range of relationships.

The most common method for achieving this is feature selection, which is generally a data preparation step in ML modelling, as it helps to select the most relevant variables when building a model. As a result, overfitting can be reduced, as less but more relevant data helps to generalise. In addition, it speeds up implementation time, improves model performance, and makes the model easier to interpret thanks to fewer variables. There are several methods for implementing this:

- Filter Method: features are selected not based on model performance, but on some statistical or information-theoretic measure (correlation, mutual information, chi-squared test, ANOVA).
- Wrapped Method: a learning algorithm is wrapped into the feature selection process, which is evaluated based on various metrics, and the method selects the best method.

- Hybrid Method: a combination of filter and wrapped methods for feature selection on large data sets [43].

We chose the embedded method for the application, as the feature selection performed during training can be efficiently queried using the model library. Figure 8 shows the most influential variables. The most important parameter is the number of reported storage errors, but as Section 2.4 shows, this is a general consequence of high storage performance rather than an influencing factor that may be important in optimising mathematical modelling. However, several other parameters can be shown to be important, which may indeed be influencing factors: the number of transport errors, the number of retailers and employees, distance values, and encoded energy certificates may also influence storage performance. The following section discusses these parameters.

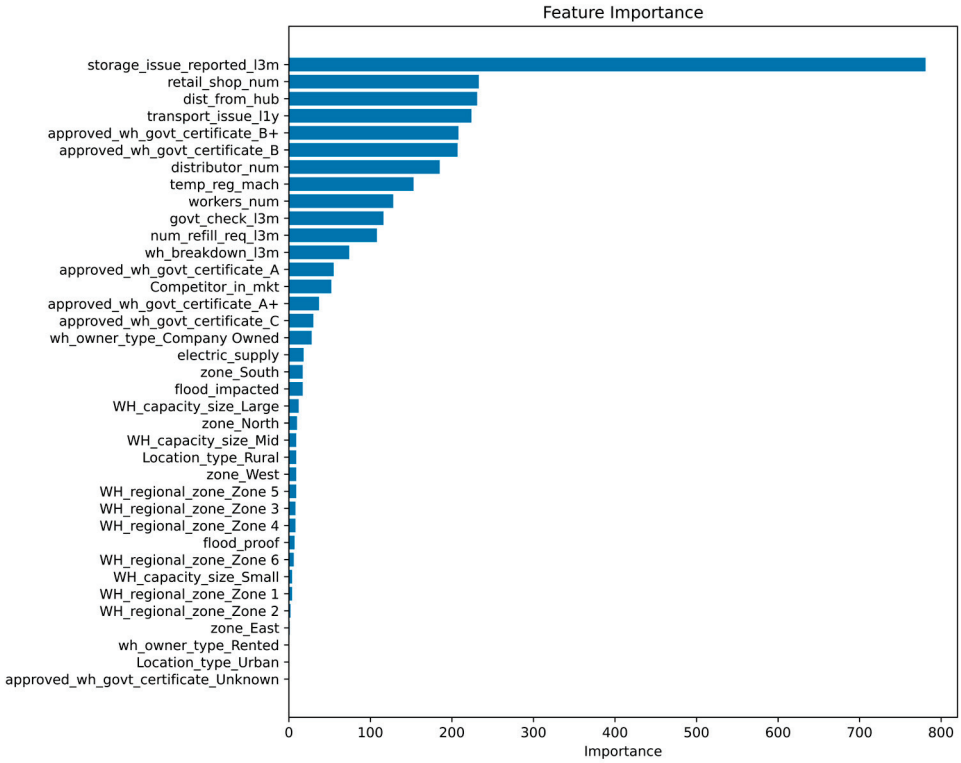


Figure 8. Feature importance based on embedding method (source: Python 3.12.11).

Another approach to extracting the importance of variables is the SHAP method, which is widely used in the interpretation of ML models, especially in the case of neural networks and GBM-based methods. The method is based on Shapley values, which are derived from game theory: the model's prediction is treated as a 'payoff' and the values are distributed among the parameters according to their influence. Thus, the summary plot shows all predictions, so each variable has a one-dimensional scatter plot. Its position on the x-axis shows how much the given parameter increases or decreases the prediction in the given observation. The colour of the point shows its characteristic value, which helps to understand which characteristic values cause positive or negative SHAP values [44]. Based on visual inspection, it can be concluded that although there are differences between the SHAP and feature importance methods, a good conclusion can be drawn about the influencing parameters in terms of pseudo-probability. Reasons why reliable conclusions can be drawn despite differences between SHAP and feature importance:

- Both methods consistently highlight overlapping key features.
- SHAP explains local, instance-level effects, while feature importance captures global trends.

- Normalization and averaging increase stability of results.
- Using two independent techniques reduces bias of a single method.
- Differences reflect complementary perspectives rather than contradictions.
- Cross-validation of results strengthens interpretability and trustworthiness.
- Robustness is enhanced because similar influential parameters appear under different assumptions.

Both methods are widely validated in ML literature for feature interpretability. After calculating the values of the two models (Figure 9), the parameter values were normalized using the MinMaxScaler method, then averaged the results of the methods, based on which the following parameters can be considered significant influencing factors:

- transport issue l1y;
- certificate;
- retail shop num;
- dist from hub;
- distributor num;
- temp reg mch;
- workers num.

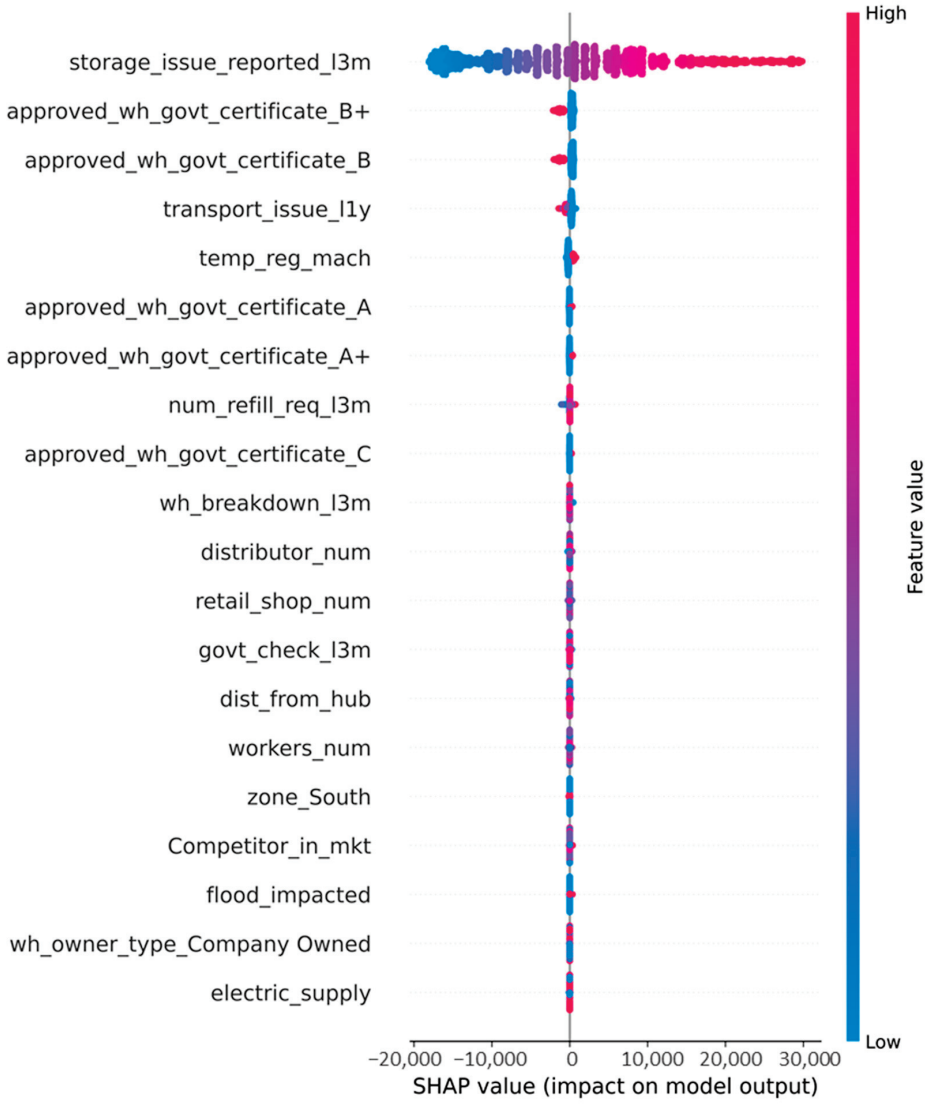


Figure 9. SHAP summary plot (source: Python 3.12.11).

3.2. Parameter Value Changes

The value of a target variable depends on various parameters, but it is possible to check what effect a given variable has on the value of the target variable. PDP is a visualisation technique that helps to understand how one or two input variables affect the predictions of a ML model on average, filtering out the effects of all other variables. The procedure helps to interpret models, supports the exploration of nonlinear relationships and the examination of feature importance.

During the research, various parameters are examined and analysed, and the level of warehouse performance achievable by changing their values is determined. The aim of the analysis is the exploration of possibilities that can be extended to real industrial problems if the costs of changing the parameters are also available. The LightGBM model was also used to create the figures.

First of all, the delivery errors that occurred in the past year are examined. In the dataset, a maximum of 5 delivery errors were recorded for a single warehouse. If all errors are reduced to zero, a 1% improvement in performance is shown by the model. However, if the number of delivery errors is reduced by only one, a 0.5% improvement can be achieved. The latter scenario is considered much more feasible, as avoiding all delivery errors requires significant investment; therefore, it is not reasonable to reduce them to zero, but avoiding a single error is realistic, and the 0.5% increase may be worthwhile. Figure 10 shows the extent of performance improvement according to the predictive model for the amount of error reduction at a given warehouse (the value '1' indicates that every storage error with a value of at least 1 is reduced by 1, while smaller values are set to 0).

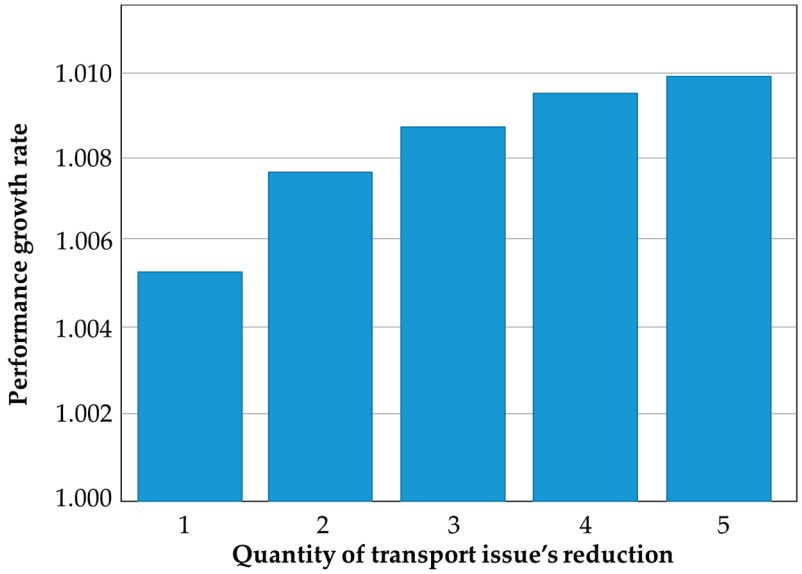


Figure 10. Transport issue reduction (source: Python 3.12.11).

Warehouse performance is related to the number of warehouse workers. However, it is also important to examine the cost of increasing the number of workers and how this relates to the surplus resulting from the increase in performance. This can be calculated precisely in the given industrial case; the research examines the determination of surplus performance using the PDP method. Figure 11 shows the rate of increase in warehouse performance if the number of employees is increased by a given multiplier. It can be concluded that doubling the number of employees would increase performance by 0.1%. Even in the absence of any other information, this cannot be considered a significant increase, but the method can be validated because feedback is obtained from a reliable model without considering the added value of the other parameters.

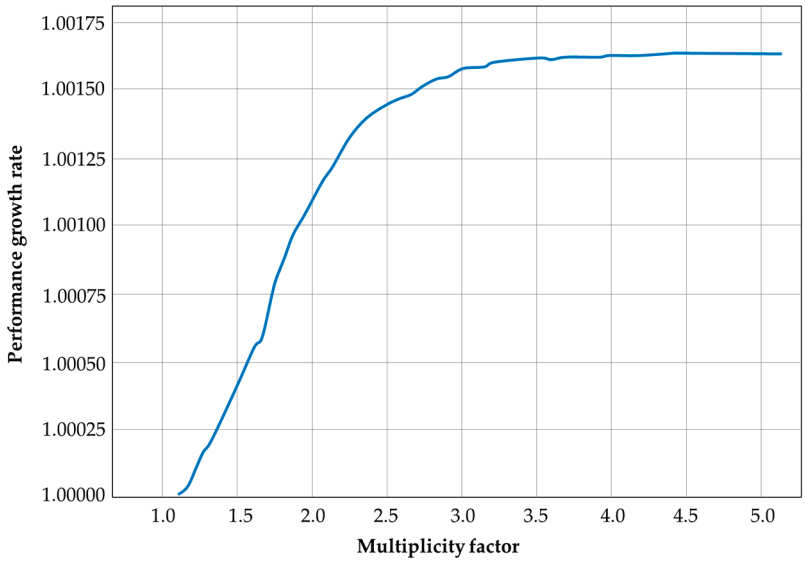


Figure 11. Workers number increasing (source: Python 3.12.11).

For some variables, reducing the parameter can increase performance, such as in the case of distance from the production unit. Figure 12 shows that if the distance between the warehouse and the production unit is reduced by a certain proportion, a minimal increase in the efficiency of warehousing activities can be observed, the extent of which is indicated by the ML model. This assumption is confirmed by several studies, which show that intelligent layout and automation can reduce travel time, resulting in faster execution and service [45].

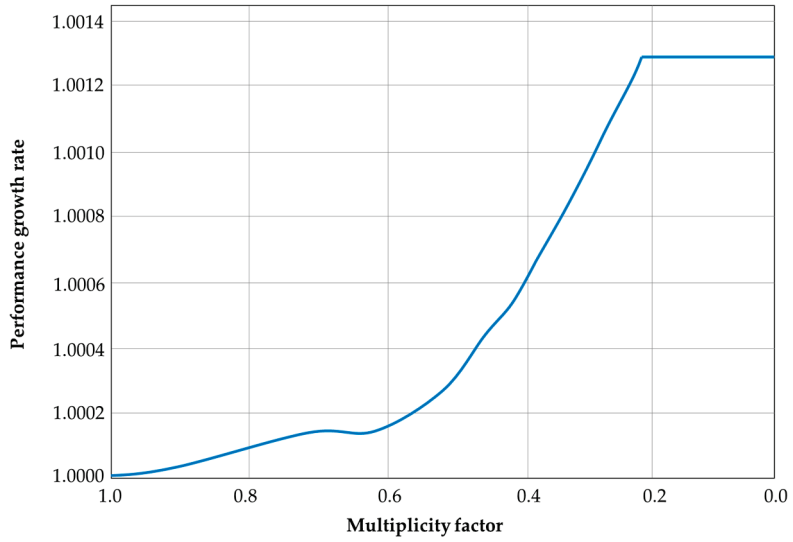


Figure 12. Decreasing distance between warehouse and production (source: Python 3.12.11).

The Partial Dependence Plot was originally developed for changing numerical values, but similar methods can also be used to analyze changes in categorical values. In this example, the energy certificate is examined, as warehouse performance was revealed by feature importance to be influenced by it. Today's modern and automated warehouses (e.g., AGVs, robots, RFID systems) require a large amount of continuous and reliable energy supply, making the quality of the energy infrastructure (e.g., substations, charging stations, energy management systems) is a critical factor for smooth operation and has a direct impact on productivity, maintenance costs, and the availability of automated systems [46].

Figure 13 shows that if a given warehouse can increase from energy level "B" or "B+" to level "A" or "A+", an efficiency increase of 9–10% can be achieved, which is by far the highest increase among the parameters examined.

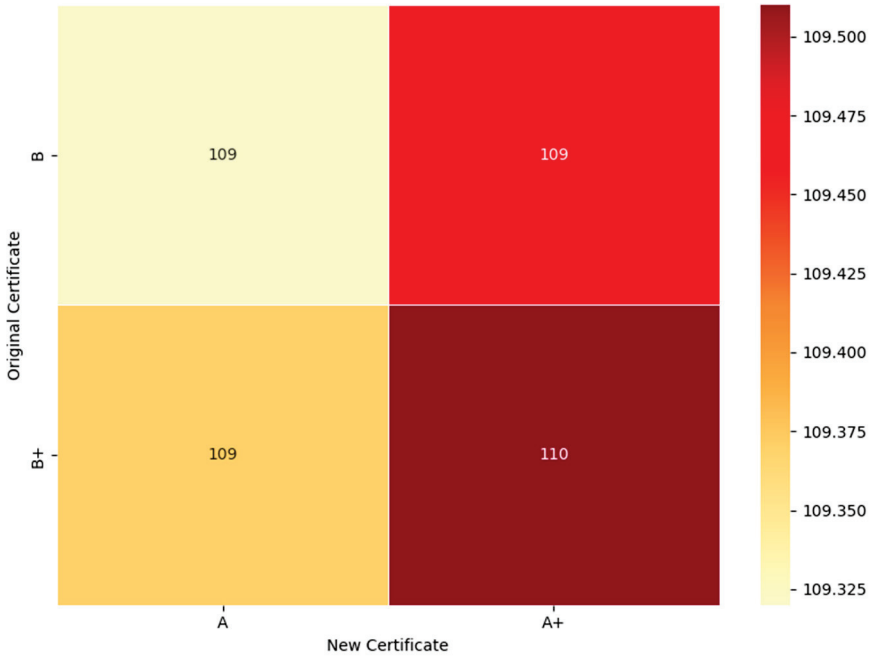


Figure 13. Matrix of the effects of energy certificate modifications (source: Python 3.12.11).

3.3. Mathematical Modelling

Mathematical modeling is a method by which a part of reality is described in abstract form using mathematical tools. During modeling, the essential characteristics of the phenomenon or system under investigation are highlighted and represented using variables, parameters, and equations.

In this research, the objective function is defined to include the parameters of transport errors presented in Section 3.2, the number of employees, the distance from the production unit, and the energy certification. Mathematical tools are applied to show the extent and trends of the changes. In addition to the maximization of performance, the costs associated with the activities must also be considered, as changing any parameter incurs a specific cost; therefore, the optimization objective function is determined according to the cost parameter.

The function describing the reduction in shipping errors refers to a saturating but increasing type of function, where the rate of performance growth decreases as more shipping problems are successfully reduced:

$$P_1(x_1) = a_1 \cdot \left(1 - e^{-b_1 x_1}\right) \tag{2}$$

where x_1 is the amount of reduction in transportation problems, P_1 is the performance increase, a_1 is the maximum growth, and b_1 is the growth rate.

For the final objective function, it is necessary to determine the cost function $C_1(x_1)$, but its trend and extent depend on the specific industrial problem, so this will not be discussed in detail in the present study. The objective function for the total parameter is as follows:

$$Z_1(x_1) = W_1 \cdot P_1(x_1) - C(x_1) \tag{3}$$

where W_1 is the weight parameter, which expresses the cost of performance increase, and C_1 is the cost function of change.

The function describing the number of employees is identical to the function describing the reduction in delivery errors, as this function is also saturating and increasing, tending towards a maximum value, so its mathematical description is also identical:

$$P_2(x_2) = a_2 \cdot \left(1 - e^{-b_2 x_2}\right) \tag{4}$$

$$Z_2(x_2) = W_2 \cdot P_2(x_2) - C(x_2), \tag{5}$$

where x_2 is the number of employees, P_2 is the performance increase, a_2 is the maximum growth, b_2 is the growth rate, W_2 is the weight parameter, which expresses the cost of performance increase, and C_2 is the cost function of change.

The distance between the storage unit and the production can also be described by a saturation function, but in reverse, as the distance decreases when performance increases:

$$P_3(x_3) = a_3 \cdot \left(1 - e^{-b_3(1-x_3)}\right) \tag{6}$$

$$Z_3(x_3) = W_3 \cdot P_3(x_3) - C(x_3), \tag{7}$$

where x_3 is the distance between storage unit and the production, P_3 is the performance increase, a_3 is the maximum growth, b_3 is the growth rate, W_3 is the weight parameter, which expresses the cost of performance increase, and C_3 is the cost function of change.

In the case of changes to the energy certificate, the increase in performance can also be written as follows:

$$P_4(x_4) = \begin{cases} 1.0932 \; if \; x_4 : B \rightarrow A \\ 1.0937 \; if \; x_4 : \; B^+ \rightarrow A \\ 1.0947 \; if \; x_4 : \; B \rightarrow A^+ \\ 1.0951 \; if \; x_4 : \; B^+ \rightarrow A^+ \end{cases} \tag{8}$$

$$Z_4(x_4) = W_4 \cdot P_4(x_4) - C(x_4), \tag{9}$$

where x_4 is the energy certificate changes, P_4 is the performance increase, W_4 is the weight parameter, which expresses the cost of performance increase, and C_4 is the cost function of change.

The total objective function of the modeling is the sum of the components presented above. This objective function must be maximized to achieve the optimum, as the components were explained based on the cost increase resulting from the performance increase, and this increase must be maximized:

$$Z = \; Z_1 + Z_2 + Z_3 + Z_4 \rightarrow max. \tag{10}$$

4. Discussion

In recent years, AI- and ML-based solutions have become increasingly popular across a range of industries. The field of logistics process optimization is broad, encompassing numerous industrial problems, and several methodologies have been developed to address them in recent decades [47,48]. Keeping pace with modern technologies, it is essential to integrate the possibilities of ML into logistics optimization.

The use of future Internet technologies, including advanced telecommunications and Internet-based solutions, represents a promising direction for improving warehousing and SC performance. These technologies enable real-time data exchange, predictive analytics, and intelligent automation, which can significantly improve the responsiveness and coordination of logistics networks. By proactively integrating these digital innovations into their warehousing activities, manufacturing companies can achieve new levels of efficiency, transparency, and scalability. This strategic alignment with emerging internet

paradigms not only supports operational optimization, but also long-term competitiveness in increasingly digitized markets and logistics [49–51].

Predictive data models are used to determine future values, as knowing these future indicators can make the planning and implementation of many logistics processes more efficient. However, these models have many other useful features that can be used for various activities. During model training, several methods (Feature Importance, SHAP, PDP) can provide feedback on how a given parameter influences the value of the target variable. Using these reports, information can be obtained about the dataset and the possibilities for influencing the target variable.

The maximization of warehouse performance is regarded as a key factor for manufacturing companies and the entire SC, as non-value-adding processes are known to significantly affect the quality of logistics processes. In this research, a fictional dataset was used to examine the applicability of ML models for optimization. The AutoML procedure was applied to select the regression method, through which the model best suited to the problem at hand was determined. The best results were obtained with the LightGBM procedure, which, with an R^2 value of 0.994, was found to be sufficiently reliable for feature selection.

The results of the research show that the Feature Importance methods of ML models can be applied with sufficient quality to understand the dataset and to determine the factors influencing the target variable. In many industrial cases, detailed information about material flow factors is not available, as these patterns are often non-trivial. The recognition of such patterns can provide several new opportunities for optimizing storage performance and improving efficiency. The Partial Dependence Plot method was used to test how changes in a given parameter affect the value of the target variable, and the diagrams created clearly show the extent of influence. Finally, mathematical modeling was employed to explain the functions describing the influencing parameters, whose curves (e.g., linear, exponential, logarithmic, etc.) were identified using the PDP method. Storage performance was defined as a cost indicator, since changes in logistics processes are costly, and the final optimal result is always determined according to cost.

Beyond the methodological results, it is important to situate the presented findings in the context of previous research on warehouse optimization. While most existing studies emphasize heuristic and metaheuristic approaches such as order-picking optimization, layout design, or reinforcement learning for dynamic slotting, the proposed regression-based approach demonstrates that interpretable ML methods can also provide valuable insights for decision support. The integration of Feature Importance and Partial Dependence Plots makes it possible not only to identify key influencing parameters but also to understand the direction and magnitude of their impact, which enhances managerial decision-making. In this sense, the described research complements existing optimization techniques by providing an alternative perspective that is strongly aligned with Industry 4.0 goals of data-driven operations and the Industry 5.0 focus on human-centric and sustainable solutions.

As a further improvement opportunity, it should be noted that testing the theoretical procedure in a real industrial environment may be crucial. It can be examined what hidden parameters the Feature Importance procedures reveal and what added value they have in actual practical optimization. In addition to maximizing storage performance, it is possible to examine the cost of changing each parameter, thus verifying the possibilities for cost-based optimization.

Despite the promising results, this study has several limitations that should be acknowledged. The analysis was conducted using a fictional and cleaned dataset rather than real industrial data, which limits the external validity and generalizability of the results. Only a subset of potentially relevant warehouse parameters was included; factors such as

real-time environmental conditions, human behavior, and SC disruptions were not modeled. The cost functions integrated into the optimization framework were conceptual rather than empirically derived, so the financial realism of the model remains to be validated. The AutoML process and LightGBM model selection were performed on a static dataset, dynamic or time-series effects were not considered, which may influence warehouse performance in practice. The study did not compare the regression-based optimization framework against other AI-driven or heuristic methods (e.g., DRL, metaheuristics) in a benchmarked environment. These limitations indicate that while the framework is theoretically robust, future research should include validation on real datasets, integration of temporal and stochastic factors, and cross-method comparison to ensure industrial applicability.

5. Conclusions

This research demonstrated that machine learning-based regression models (particularly the Light Gradient Boosting Machine (LightGBM) algorithm identified through AutoML) can effectively uncover hidden performance patterns in warehouse operations. The combined use of Feature Importance and Partial Dependence Plot (PDP) analyses enabled the detection of the most influential parameters and the visualization of their effects on the target variable. By translating these insights into mathematical relationships, the study introduced a novel, data-driven approach for constructing optimization-oriented objective functions. This framework enhances both predictive accuracy and interpretability, offering a transparent, cost-based perspective that supports more efficient and informed decision-making in warehouse management.

Despite these promising results, several limitations must be acknowledged. The analysis focused on a limited range of parameters, whereas real-world warehouse performance is likely affected by a broader set of operational, environmental, and human factors. In addition, the cost functions were defined conceptually rather than empirically, which may constrain the immediate industrial applicability of the proposed optimization model.

Future research should address these limitations by validating the methodology with real industrial datasets and extending the scope of influencing variables. Particular attention should be given to quantifying cost structures and embedding them into the optimization process to ensure practical relevance. Moreover, integrating regression-based interpretability with reinforcement learning or simulation-based methods could open new directions for adaptive, real-time warehouse optimization. Finally, cross-industry comparative studies are recommended to evaluate the robustness, scalability, and generalizability of the proposed framework under diverse operational conditions.

Author Contributions: Conceptualization, Á.F. and T.B.; methodology, T.B.; software, Á.F.; validation, Á.F.; formal analysis, Á.F. and T.B.; investigation, Á.F. and T.B.; resources, Á.F.; data curation, Á.F.; writing—original draft preparation, Á.F. and T.B.; writing—review and editing, Á.F. and T.B.; visualization, Á.F.; supervision, T.B.; project administration, T.B.; funding acquisition, T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are unavailable due to privacy restrictions.

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACO	Ant Colony Optimization
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
ANOVA	Analysis of Variance
API	Application Programming Interface
AutoML	Automated Machine Learning
COI	Cube Per Order Index
DES	Discrete Event Simulation
DRL	Deep Reinforcement Learning
DSLAP	Dynamic Storage Location Assignment Problem
GA	Genetic Algorithms
GBM	Gradient Boosting Machine
IoT	Internet of Things
IT	Information Technology
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean Squared Error
OOS	Order Oriented Slotting
PDP	Partial Dependence Plot
RFID	Radio Frequency Identification
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Logarithmic Error
SC	Supply Chain
SHAP	SHapley Additive exPlanations
SKU	Stock Keeping Unit

References

1. Negi, P.; Pathani, A.; Bhatt, B.C.; Swami, S.; Singh, R.; Gehlot, A.; Thakur, A.K.; Gupta, L.R.; Priyadarshi, N.; Twala, B.; et al. Integration of Industry 4.0 Technologies in Fire and Safety Management. *Fire* **2024**, *7*, 335. [CrossRef]
2. Li, K. Optimizing Warehouse Logistics Scheduling Strategy Using Soft Computing and Advanced Machine Learning Techniques. *Soft Comput.* **2023**, *27*, 18077–18092. [CrossRef]
3. Savaglio, C.; Mazzei, P.; Fortino, G. Edge Intelligence for Industrial IoT: Opportunities and Limitations. *Procedia Comput. Sci.* **2024**, *232*, 397–405. [CrossRef]
4. Hu, Y.; Jia, Q.; Yao, Y.; Lee, Y.; Lee, M.; Wang, C.; Zhou, X.; Xie, R.; Yu, F.R. Industrial Internet of Things Intelligence Empowering Smart Manufacturing: A Literature Review. *IEEE Internet Things J.* **2024**, *11*, 19143–19167. [CrossRef]
5. Hosseini, M.; Chalil Madathil, S.; Khasawneh, M.T. Reinforcement Learning-Based Simulation Optimization for an Integrated Manufacturing-Warehouse System: A Two-Stage Approach. *Expert Syst. Appl.* **2025**, *290*, 128259. [CrossRef]
6. de Koster, R.; Le-Duc, T.; Roodbergen, K.J. Design and Control of Warehouse Order Picking: A Literature Review. *Eur. J. Oper. Res.* **2007**, *182*, 481–501. [CrossRef]
7. Roodbergen, K.J. *Layout and Routing Methods for Warehouses*, 1st ed.; Erasmus University Rotterdam: Rotterdam, The Netherlands, 2001.
8. Small Parts Order Picking: Design and Operation. Available online: https://www2.isye.gatech.edu/~mgoetsch/cali/Logistics%20Tutorial/order/article.htm (accessed on 12 August 2025).
9. Hausman, W.H.; Schwarz, L.B.; Graves, S.C. Optimal storage assignment in automatic warehousing systems. *Manag. Sci.* **1976**, *22*, 629–638. [CrossRef]
10. Schuur, P.C. *The Cube Per Order Index Slotting Strategy, How Bad Can It Be?* University of Twente: Enschede, The Netherlands, 2014.
11. Mantel, R.J.; Schuur, P.C.; Heragu, S.S. Order oriented slotting: A new assignment strategy for warehouses. *Eur. J. Ind. Eng.* **2007**, *1*, 301. [CrossRef]
12. De Koster, M.B.M.; Van der Poort, E.S.; Wolters, M. Efficient orderbatching methods in warehouses. *Int. J. Prod. Res.* **1999**, *37*, 1479–1504. [CrossRef]

13. Tsige, M. Improving Order-Picking Efficiency Via Storage Assignments Strategies. Master's Thesis, University of Twente, Enschede, The Netherlands, 2013.
14. Waubert de Puiseau, C.; Nanfack, D.T.; Tercan, H.; Löbbert-Plattfaut, J.; Meisen, T. Dynamic Storage Location Assignment in Warehouses Using Deep Reinforcement Learning. *Technologies* **2022**, *10*, 129. [CrossRef]
15. Seyedan, M.; Mafakheri, F. Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *J. Big Data* **2020**, *7*, 53. [CrossRef]
16. Aamer, A.; Eka Yani, L.; Alan Priyatna, I. Data Analytics in the Supply Chain Management: Review of Machine Learning Applications in Demand Forecasting. *Oper. Supply Chain. Manag. Int. J.* **2020**, *14*, 1–13. [CrossRef]
17. Falatouri, T.; Darbanian, F.; Brandtner, P.; Udokwu, C. Predictive Analytics for Demand Forecasting—A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Comput. Sci.* **2020**, *200*, 993–1003. [CrossRef]
18. Douaioui, K.; Oucheikh, R.; Benmoussa, O.; Mabrouki, C. Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review. *Appl. Syst. Innov.* **2024**, *7*, 93. [CrossRef]
19. Omprakash, M.K. Optimizing Demand Forecasting and Inventory Management with AI in Automotive Industry. Master's Thesis, Lappeenranta–Lahti University of Technology LUT, Lappeenranta, Finland, 2024.
20. Umoren, J.; Utomi, E.; Adukpo, T. AI-powered Predictive Models for U.S. Healthcare Supply Chains: Creating AI Models to Forecast and Optimize Supply Chain. *IJMR* **2025**, *11*, 2455–3662.
21. Srinivas, A. AI-Driven Demand Forecasting in Enterprise Retail Systems: Leveraging Predictive Analytics for Enhanced Supply Chain. *Int. J. Sci. Technol.* **2025**, *16*, 1–22. [CrossRef]
22. Cergibozan, Ç.; Tasan, A.S. Genetic algorithm based approaches to solve the order batching problem and a case study in a distribution center. *J. Intell. Manuf.* **2022**, *33*, 137–149. [CrossRef]
23. Azadnia, A.H.; Taheri, S.; Ghadimi, P.; Mat Saman, M.Z.; Wong, K.Y. Order Batching in Warehouses by Minimizing Total Tardiness: A Hybrid Approach of Weighted Association Rule Mining and Genetic Algorithms. *Sci. World J.* **2013**, *2013*, 246578. [CrossRef]
24. de Koster, R.; van der Poort, E.; Roodbergen, K.J. When to apply optimal or heuristic routing of orderpickers. In *Advances in Distribution Logistics*, 1st ed.; Fleischmann, B., van Nunen, J.A.E.E., Speranza, M.G., Stähly, P., Eds.; Springer: Berlin, Germany, 1998; Volume 460, pp. 375–401. [CrossRef]
25. Cano, J.A.; Correa-Espinal, A.A.; Gómez-Montoya, R.A. An Evaluation of Picking Routing Policies to Improve Warehouse Efficiency. *Int. J. Ind. Eng. Manag.* **2017**, *8*, 229–238. [CrossRef]
26. Tufano, A.; Accorsi, R.; Manzini, R. A machine learning approach for predictive warehouse design. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 2369–2392. [CrossRef]
27. Wang, Y.; Vasan, G.; Mahmood, R. Real-Time Reinforcement Learning for Vision-Based Robotics Utilizing Local and Remote Computers. *arXiv* **2022**, arXiv:2210.02317. [CrossRef]
28. Peyas, I.S.; Hasan, Z.; Tushar, M.R.R.; Musabbir, A.; Azni, R.M.; Siddique, S. *Autonomous Warehouse Robot using Deep Q-Learning*; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]
29. Vinod Chandra, S.S.; Anand, H.S. Nature inspired meta heuristic algorithms for optimization problems. *Computing* **2022**, *104*, 251–269. [CrossRef]
30. Discrete Event Simulation (DES). Available online: https://www.ebsco.com/research-starters/mathematics/discrete-event-simulation-des (accessed on 1 September 2025).
31. Georgescu, I. The early days of Monte Carlo methods. *Nat. Rev. Phys.* **2023**, *5*, 372. [CrossRef]
32. Antony, J. Six Sigma for service processes. *Bus. Process Manag. J.* **2006**, *12*, 234–248. [CrossRef]
33. Yao, J.F.; Yang, Y.; Wang, X.C.; Zhang, X.P. Systematic review of digital twin technology and applications. *Vis. Comput. Ind. Biomed.* **2023**, *6*, 10. [CrossRef]
34. Dubey, V.; Kumari, P.; Patel, K.; Singh, S.; Shrivastava, S. Amalgamation of Optimization Algorithms with IoT Applications. In *Sustainable Development in Industry and Society 5.0*; IGI Global: Hershey, PA, USA, 2024; pp. 176–204. [CrossRef]
35. Wen, X.; Liao, J.; Niu, Q.; Shen, N.; Bao, Y. Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. *Sci. Rep.* **2024**, *14*, 13720. [CrossRef]
36. Edozie, E.; Shuaibu, A.N.; John, U.K.; Sadiq, B.O. Comprehensive review of recent developments in visual object detection based on deep learning. *Artif. Intell. Rev.* **2025**, *58*, 277. [CrossRef]
37. Ma, C.; Li, A.; Du, Y.; Dong, H.; Yang, Y. Efficient and scalable reinforcement learning for large-scale network control. *Nat. Mach. Intell.* **2024**, *6*, 1006–1020. [CrossRef]
38. Supply Chain Optimization for a FMCG Company. Available online: https://www.kaggle.com/datasets/suraj9727/supply-chain-optimization-for-a-fmcg-company/data (accessed on 20 July 2025).
39. He, X.; Zhao, K.; Chu, X. AutoML: A survey of the state-of-the-art. *Knowl.-Based Syst.* **2021**, *212*, 106622. [CrossRef]
40. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]

41. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
42. LightGBM Documentation. Available online: https://lightgbm.readthedocs.io/en/latest/Features.html# (accessed on 16 August 2025).
43. Theng, D.; Bhoyar, K.K. Feature selection techniques for machine learning: A survey of more than two decades of research. *Knowl. Inf. Syst.* **2024**, *66*, 1575–1637. [CrossRef]
44. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 1013–1026. [CrossRef]
45. Reducing Picker Travel Time: Enhancing Warehouse Efficiency with Automation and Smart Slotting. Available online: https://www.logiwa.com/blog/reducing-picker-travel-time-to-enhance-warehouse-efficiency (accessed on 17 August 2025).
46. Transforming Logistics: The Rise of High-Power Warehouses and Smart Energy Systems. Available online: https://blog.intellimeter.com/transforming-logistics-the-rise-of-high-power-warehouse-and-smart-energy-systems (accessed on 17 August 2025).
47. Korponai, J.; Bányai, Á.; Illés, B. The Effect of the Safety Stock on the Occurrence Probability of the Stock Shortage. *Manag. Prod. Eng. Rev.* **2017**, *8*, 69–77. [CrossRef]
48. Borodavko, B.; Illés, B.; Bányai, Á. Role of Artificial Intelligence in Supply Chain. *Acad. J. Manuf. Eng.* **2021**, *19*, 75–79.
49. Veres, P. A Comparison of the Black Hole Algorithm against Conventional Training Strategies for Neural Networks. *Mathematics* **2025**, *13*, 2416. [CrossRef]
50. Veres, P. ML and Statistics-Driven Route Planning: Effective Solutions without Maps. *Logistics* **2025**, *9*, 124. [CrossRef]
51. Veres, P. The Opportunities and Possibilities of Artificial Intelligence in Logistic Systems: Principles and Techniques. In *Advances in Digital Logistics, Logistics and Sustainability*; Tamás, P., Bányai, T., Telek, P., Cservenák, Á., Eds.; CECOL 2024. Lecture Notes in Logistics; Springer: Cham, Switzerland, 2024. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

Toward the Theoretical Foundations of Industry 6.0: A Framework for AI-Driven Decentralized Manufacturing Control

Andrés Fernández-Miguel [1,2], Susana Ortíz-Marcos [3], Mariano Jiménez-Calzado [3], Alfonso P. Fernández del Hoyo [1], Fernando E. García-Muiña [2] and Davide Settembre-Blundo [1,4,*]

[1] Faculty of Economics and Business Administration (ICADE), Comillas Pontifical University, 28015 Madrid, Spain; afmiguel@icade.comillas.edu (A.F.-M.); fdelhoyo@icade.comillas.edu (A.P.F.d.H.)
[2] Department of Business Administration (ADO), Rey Juan Carlos University, 28933 Madrid, Spain; fernando.muina@urjc.es
[3] School of Engineering (ICAI), Comillas Pontifical University, 28015 Madrid, Spain; sortiz@iit.comillas.edu (S.O.-M.); mjimenez@icai.comillas.edu (M.J.-C.)
[4] Innovability Unit, Gresmalt Group, 41049 Sassuolo, Italy
* Correspondence: dsettembre@comillas.edu

Abstract

This study advances toward establishing the theoretical foundations of Industry 6.0 by developing a comprehensive framework that integrates artificial intelligence (AI), decentralized control systems, and cyber–physical production environments for intelligent, sustainable, and adaptive manufacturing. The research employs a tri-modal methodology (deductive, inductive, and abductive reasoning) to construct a theoretical architecture grounded in five interdependent constructs: advanced technology integration, decentralized organizational structures, mass customization and sustainability strategies, cultural transformation, and innovation enhancement. Unlike prior conceptualizations of Industry 6.0, the proposed framework explicitly emphasizes the cyclical feedback between innovation and organizational design, as well as the role of cultural transformation as a binding element across technological, organizational, and strategic domains. The resulting framework demonstrates that AI-driven decentralized control systems constitute the cornerstone of Industry 6.0, enabling autonomous real-time decision-making, predictive zero-defect manufacturing, and strategic organizational agility through distributed intelligent control architectures. This work contributes foundational theory and actionable guidance for transitioning from centralized control paradigms to AI-driven distributed intelligent manufacturing control systems, establishing a conceptual foundation for the emerging Industry 6.0 paradigm.

Keywords: Industry 6.0; theoretical foundations; AI-driven control systems; decentralized manufacturing; cyber-physical systems; intelligent automation; Industrial Internet of Things; smart manufacturing; sustainable production; mass customization; manufacturing paradigm

1. Introduction

Establishing the theoretical foundations of Industry 6.0 requires understanding how the convergence of artificial intelligence (AI) and cyber–physical control systems is fundamentally transforming industrial manufacturing, driving evolution beyond established paradigms toward unprecedented levels of intelligent automation and sustainable production. Recent decades have witnessed accelerated industrial transformation through various

evolutionary stages, culminating in the emerging concept of Industry 6.0 [1]. Following the digital transformation of Industry 4.0 [2], characterized by cyber–physical systems integration and intelligent automation, and the human-centric approach of Industry 5.0 [3], which emphasizes collaborative human–machine interaction and sustainability-oriented manufacturing, Industry 6.0 represents the next evolutionary paradigm that fully integrates AI-driven control systems with sustainable production environments [4]. Importantly, Industry 5.0 should not be understood as a concluded paradigm but as an ongoing research and policy agenda with early industrial implementations. In this respect, our framework positions Industry 6.0 as an emerging and complementary paradigm that extends, rather than replaces, the principles of Industry 5.0. Industry 6.0 represents the cognitive manufacturing paradigm where knowledge is transformed into autonomous "thought-like" decision-making capabilities through AI-driven systems that exhibit cognitive functions such as learning, reasoning, and adaptive problem solving. Unlike the data-centric approach of Industry 4.0 or the knowledge-collaborative focus of Industry 5.0, Industry 6.0 establishes cognitive manufacturing systems capable of autonomous strategic thinking, operating within cyclical feedback loops between technological cognition, organizational intelligence, strategic adaptation, cultural evolution, and continuous innovation. This cognitive dimension, combined with the cyclical interdependence of constructs rather than linear technological advancement, constitutes the paradigm's distinguishing theoretical foundation and primary contribution to manufacturing theory.

Originality and Contribution of the Framework

The originality of the proposed framework lies in how it advances beyond existing Industry 4.0 and 5.0 conceptualizations. Whereas Industry 4.0 has largely been framed around cyber–physical integration and centralized digital efficiency, and Industry 5.0 has focused on human–machine collaboration and sustainability awareness, our framework introduces two key contributions. First, it highlights the cyclical feedback loops across technological, organizational, strategic, cultural, and innovation constructs, moving away from linear or siloed representations that dominate previous models. Second, it identifies cultural transformation as a binding element that systematically connects all dimensions, rather than treating culture as a background condition or maturity stage.

These contributions position Industry 6.0 not as a replacement but as a theoretical extension of previous paradigms, integrating dispersed elements of the literature [1,3,5] into a coherent and systemic architecture. This integrated approach, illustrated in Table 1 and further developed in the conceptual models, underscores the novelty of our contribution to both academic debate and managerial practice.

Industry 6.0 can be understood as a cognitive manufacturing paradigm in which AI-driven systems and distributed control architectures transform knowledge into autonomous and adaptive decision-making capabilities. The model is cyclical and interdependent: advanced technology integration (AI and IIoT) reconfigures organizational structures (decentralization), enables strategies that combine mass customization and sustainability, requires continuous cultural transformation, and fosters proactive innovation. Innovation, in turn, feeds back into technology and organizational structures, closing the loop.

Unlike prior conceptualizations that emphasize primarily techno-centric [6] or human-centric [7] and sustainability-oriented [8] perspectives, our definition explicitly stresses the systemic feedback loops and the binding role of cultural transformation as distinctive contributions. This working definition also provides the interpretative lens for Table 1, which summarizes how Industry 6.0 extends and integrates the main epistemological, organizational, and strategic features of Industry 4.0 and 5.0.

Table 1. Distinctive Characteristics Across Industrial Paradigms.

Dimension	Industry 4.0	Industry 5.0	Industry 6.0
Epistemological Foundation	Data collection and processing	Information to knowledge transformation	Knowledge to cognitive "thought"
Primary Focus	Operational efficiency through digitalization	Human–machine collaborative knowledge	Cognitive manufacturing systems
Decision-Making	Data-driven with human oversight	Knowledge-based human–machine collaboration	Autonomous cognitive decision-making
Organizational Structure	Traditional hierarchies with digital tools	Enhanced collaboration within existing structures	Cognitively enabled decentralized architectures
Strategic Orientation	Cost reduction and efficiency	Customization with sustainability awareness	Cognitive mass customization–sustainability integration
Cultural Foundation	Process optimization culture	Human-centric collaborative values	Continuous cognitive transformation
Innovation Pattern	Technology-driven improvements	Collaborative knowledge innovation	Autonomous cognitive innovation loops
Theoretical Novelty	CPS and IoT integration	Human-centricity in knowledge work	Cyclical cognitive feedback between all constructs

This transformation toward intelligent control systems and decentralized manufacturing architectures [9] aligns with the growing need for organizational resilience, adaptation to rapid change, and flexible responses to external disruptions, as demonstrated by the COVID-19 pandemic [10]. The pandemic served as a catalyst for Industry 6.0 advancement, exposing vulnerabilities in traditional manufacturing control systems and highlighting the critical importance of cyber–physical systems capable of autonomous decision-making [11]. The imperative to respond rapidly to supply chain disruptions and dynamic market conditions has accelerated the integration of artificial intelligence and Industrial Internet of Things (IIoT) technologies to create more resilient and adaptive manufacturing ecosystems. These enabling technologies, foundational in Industry 4.0 [12] and 5.0 [13], have evolved in Industry 6.0 to enable predictive analytics, autonomous optimization, and real-time adaptability, establishing a robust foundation for sustainable long-term manufacturing resilience [14].

This research addresses the growing need for intelligent and adaptive control architectures in modern manufacturing environments. Building on the technological foundations of Industry 4.0 [15] and the human-centric principles of Industry 5.0 [16], the proposed framework explores how artificial intelligence, and decentralized decision-making can converge to create resilient, sustainable, and highly customized production systems. By focusing on distributed control, predictive optimization, and real-time adaptability, this study positions Industry 6.0 as an evolutionary step that integrates advanced technologies with organizational and cultural transformation, offering a structured approach to guide both academic inquiry and industrial practice.

Industry 6.0 fundamentally redefines manufacturing paradigms, moving beyond the standardized mass production of the 20th century and transcending the limited customization capabilities of Industry 4.0 and 5.0 [5]. The paradigm prioritizes mass customization integrated with sustainability principles, compelling organizations to develop strategies that seamlessly combine operational flexibility with environmental responsibility. Contemporary consumers, increasingly conscious of environmental impacts, demand personalized products that align with sustainability imperatives [17]. Manufacturing organizations must

consequently adopt intelligent control systems that optimize efficiency and customization while minimizing environmental footprint, making circular production and responsible resource management integral components of smart manufacturing strategies [18].

A defining characteristic of Industry 6.0 is its commitment to achieving zero-defect manufacturing environments through predictive analytics and AI-enabled quality control systems, extending and revolutionizing the quality objectives of previous industrial paradigms [19]. This ambitious goal necessitates advanced integration of monitoring technologies, cyber–physical systems, and predictive algorithms that proactively eliminate defects while optimizing product quality and sustainability outcomes [20]. The zero-defect approach compels organizations to fundamentally restructure both production processes and organizational architectures, creating more dynamic, responsive systems focused on continuous improvement and autonomous optimization. This comprehensive transformation represents a strategic paradigm shift that embraces integrated quality and sustainability management through AI-driven control systems, positioning Industry 6.0 as the cornerstone of modern intelligent manufacturing strategies [21].

Despite extensive literature documenting Industry 4.0 and 5.0 impacts, three critical theoretical gaps persist that justify the cognitive manufacturing paradigm of Industry 6.0. First, existing frameworks treat technological, organizational, strategic, and cultural dimensions as independent variables rather than cyclically interdependent constructs enabled by cognitive systems. Second, while AI and cyber–physical systems have been extensively studied as data processors (4.0) and knowledge enablers (5.0), their evolution toward cognitive "thought-like" capabilities that fundamentally reconfigure organizational DNA remains theoretically underdeveloped [22]. Third, the integration of mass customization and sustainability as mutually reinforcing strategic imperatives, enabled by cognitive decision-making systems rather than competing human-managed objectives, lacks a theoretical foundation. These gaps necessitate a comprehensive cognitive manufacturing framework that explicates the cyclical relationships between constructs, positioning cultural transformation as the binding element that enables continuous cognitive organizational evolution [23]. Additionally, although zero-defect concepts have been discussed in manufacturing literature, limited research has investigated the long-term organizational and cultural implications of implementing AI-driven zero-defect environments on strategic transformation and innovation capabilities [24].

The lack of a standardized definition of Industry 6.0 reinforces the motivation for this study. Existing works offer heterogeneous and sometimes conflicting descriptions [1,13,20], with some emphasizing techno-centric perspectives and others human-centric or sustainability-driven orientations. In this fragmented landscape, developing a coherent theoretical foundation becomes essential for guiding both academic debate and managerial practice. Our framework does not attempt to impose a definitive definition of Industry 6.0; rather, it integrates dispersed elements into a systemic and cyclical architecture that clarifies how technologies, strategies, and cultural drivers interact to form a distinct industrial paradigm.

Given these gaps, the following research questions arise:

- RQ1: How does Industry 6.0 influence the structure and strategy of industrial organizations?
- RQ2: What organizational structures are best suited to support the decentralized decision-making and integration of advanced technologies foreseen by Industry 6.0?
- RQ3: How can organizations develop strategies that effectively balance mass customization and sustainability within the Industry 6.0 model?
- RQ4: What are the cultural and strategic implications of a zero-fault environment, and how does it affect continuous innovation and incremental improvement in industrial organizations?

2. Methodology

This research adopts a systematic and epistemologically rigorous methodology, integrating deductive [25], inductive [26], and abductive [27] reasoning across distinct phases to build a comprehensive theoretical framework for Industry 6.0.

In the deductive phase, we developed research propositions from an extensive review of literature covering Industry 4.0 and 5.0 paradigms, as well as emerging contributions on Industry 6.0 [28]. Key theoretical constructs and principles were identified to shape initial assumptions about organizational, cultural, and technological changes.

The inductive phase involved analyzing systematic reviews and recent empirical studies to detect gaps in the existing knowledge base [29]. We clustered these gaps into thematic areas such as (i) cultural implications of zero-defect manufacturing, (ii) the integration of sustainability into control systems, and (iii) decentralized decision-making in AI-driven environments. From these clusters, additional propositions were formulated to address overlooked or underexplored aspects.

Finally, the abductive phase connected these constructs into a unified theoretical model [30]. Abduction was operationalized through inference-to-the-best-explanation: iterative cycles of reasoning linked empirical anomalies and conceptual gaps to new theoretical relationships. For instance, the absence of explicit cultural considerations in zero-defect literature was abductively explained by integrating a construct on cultural transformation. This process strengthened the internal logic of the framework by aligning theoretical insights with emerging industrial practices.

Overall, this tri-modal reasoning approach ensures that the proposed model is both conceptually rigorous and grounded in empirical relevance. Deduction provided theoretical grounding, induction revealed knowledge gaps, and abduction allowed for the creative integration of constructs into a holistic model of Industry 6.0.

3. Industry 4.0 to 6.0: Theoretical Foundations and Strategic Adaptations

3.1. Moving from Industry 4.0 to 6.0

The progression from Industry 4.0 to Industry 6.0 reflects profound changes in technological integration, organizational design, and strategic orientation. Industry 4.0 introduced cyber–physical systems (CPSs) and the Internet of Things (IoT), creating a foundation for digital automation and efficiency gains [31]. However, decision-making largely remained centralized, limiting flexibility and responsiveness [32]. As a result, the potential benefits of these new technologies for agility and adaptability were not fully realized [33]. Industry 5.0 expanded this paradigm by emphasizing human–machine collaboration, personalization, and initial steps toward sustainability [34]. Yet, structural change was modest, with organizations maintaining traditional hierarchies that restricted agility [35,36]. Based on these findings and comparisons, the following considerations can be made:

- P1: *The integration of cyber–physical systems and IoT in Industry 4.0 laid the foundation for digital automation but maintained centralized organizational structures and strategies focused on efficiency [31,32].*
- P2: *The transition to Industry 5.0 introduced human–machine collaboration, improved personalization capabilities, and began to incorporate sustainability, but without significant changes to organizational structures [34,36].*

By contrast, Industry 6.0 [5] builds on these advances with the integration of artificial intelligence (AI) and the Industrial Internet of Things (IIoT) to enable autonomous systems and decentralized decision-making [37]. Business strategy is reoriented around the dual imperatives of mass customization and sustainability [38,39]. Based on these observations and comparisons, the following propositions can be formulated:

- P3: *The evolution towards Industry 6.0 involves significant technological advances with AI and IIoT enabling autonomous systems, requiring decentralized organizational structures to fully exploit these technologies [5,37].*
- P4: *In Industry 6.0, the business strategy focuses on mass customization and sustainability as central objectives, building on the personalization of Industry 5.0 and addressing the limitations of efficiency-focused strategies of Industry 4.0 [38,39].*

Crucially, Industry 6.0 requires a cultural transformation that embeds continuous improvement and sustainability as organizational values [40]. In our framework, cultural transformation is not addressed as a generic change initiative but as a structural binding construct that systematically connects technological integration, decentralized organizational forms, strategic orientation, and innovation. Unlike Industry 4.0 and 5.0 maturity models, where culture is treated as an enabling factor or peripheral dimension [41], Industry 6.0 positions cultural transformation as both an outcome of cognitive manufacturing (e.g., zero-defect learning environments, sustainability-driven values) and a precondition for its consolidation. This cultural dimension, functioning as the connective tissue of the paradigm, also enables organizations to leverage advanced analytics and autonomous systems to drive proactive, data-driven innovation [42]. Based on the above findings and comparisons, we can state the following:

- P5: *Cultural change is essential in Industry 6.0, fostering a culture of continuous improvement and sustainability values, moving beyond the process optimization culture of Industry 4.0 and the human-centric focus of Industry 5.0 [40,43].*
- P6: *Innovation in Industry 6.0 will be proactive and data-driven, enabled by advanced analytics and autonomous systems, leading to continuous improvement in strategy and technology adoption [42].*

This set of propositions highlights where our framework advances beyond existing conceptualizations: the cyclical feedback between innovation (P6 $\rightarrow$ C5) and organizational/technological design (C1, C2) and the role of cultural transformation (P5 $\rightarrow$ C4) as a structural connector. These features constitute the main theoretical novelty of our approach.

3.2. Technological Integration in Industry 6.0

Technological integration is a core element of Industry 6.0, characterized by the implementation of advanced technologies such as artificial intelligence (AI) [44] and the Industrial Internet of Things (IIoT) [45], alongside a strong commitment to sustainability across the entire production cycle. These systems drive improvements in operational efficiency and expand flexibility, enabling organizations to respond quickly to market shifts and environmental challenges [46].

Industry 4.0 focused on digitization and automation using cyber–physical systems and IoT, aiming to create smart factories that increase productivity and efficiency [47]. Although AI and IoT were introduced to automate workflows and enable data collection, human decision-making remained central, and sustainability was not yet a priority [48]. With Industry 5.0, there was a clear convergence between technological sophistication and human critical thinking [49]. This development led to more customized products [50], and while sustainability entered the conversation, it was not fully integrated into operational frameworks [51].

Industry 6.0 marks a decisive shift. It embeds autonomous and intelligent resource management into a holistic, sustainability-focused approach [52]. Artificial intelligence enables systems to learn from data and make independent decisions, enhancing precision and reducing the need for human intervention [53]. The IIoT supports real-time data acquisition, allowing equipment within Industry 6.0 environments to act autonomously [54].

Continuous monitoring helps organizations track operations, identify areas for improvement, and anticipate challenges. This methodology ensures efficient and agile solutions are prioritized [55].

Sustainability integration within the Industry 6.0 framework represents a significant progression from prior practices [56]. A defining feature of Industry 6.0 is its explicit attention to the environmental consequences of technology adoption [57]. Utilizing AI and IIoT, organizations monitor and optimize energy use, resource allocation, and waste reduction in real time [58]. This supports alignment with ecological goals, encourages more efficient resource management, helps reduce carbon footprints, and strengthens competitive positioning [59]. For example, lowering energy consumption per unit output not only boosts efficiency but also enhances the corporate image among consumers and stakeholders who increasingly value sustainability [60]. Drawing from these analyses and comparisons, the following propositions can be stated:

- P7: *Compared to previous paradigms, the combination of advanced technologies such as AI and IIoT in Industry 6.0 promotes predictive optimization and intelligent automation, significantly improving the operational efficiency and flexibility of industrial organizations through autonomous decision-making systems [44,45].*
- P8: *Integrating sustainable practices through AI and IIoT in Industry 6.0 enables organizations to balance mass customization with environmental responsibility, strengthening strategic positioning and corporate reputation beyond what was achieved in Industry 4.0 and Industry 5.0 [56,58].*

Siemens' Amberg factor (Siemens Digital Enterprise, Available online: https://www.siemens.com/global/en/company/stories/industry/electronics-digitalenterprise-futuretechnologies.html; accessed on 29 August 2025) demonstrates cognitive manufacturing, where AI autonomously adjusts production based on real-time quality predictions. By analyzing control point data, the system anticipates defects and proactively modifies processes, enabling strategic decisions about scheduling, resources, and quality without human input.

3.3. Organizational Structure Adaptations

In Industry 6.0, adapting organizational structures is crucial. Companies need flexible frameworks that enable rapid responses to market dynamics and evolving consumer demands. Within a connected manufacturing environment, decentralized decision-making and adaptable organizational forms are essential for competitiveness [61]. In Industry 4.0, firms mainly retained conventional hierarchies, relying on automation and digitalization to enhance efficiency. However, decision-making mostly remained centralized [62]. While these systems improved operations, they were not designed to address swift market transitions [63]. Industry 5.0 brought greater collaboration between humans and machines, creating a more adaptive work context, yet control often stayed centralized, limiting true organizational agility [64].

Industry 6.0 advances these developments by favoring decentralization as a structural strategy for responding more efficiently to market fluctuations. This enables organizations to react more promptly to changing conditions, as localized decision-making shortens response times [65]. With the integration of AI and IIoT, business units can access real-time data, allowing autonomous decisions independent of central management and enhancing overall system responsiveness [66].

To align with shifting consumer expectations and the growing importance of sustainable practices, companies must further restructure themselves in Industry 6.0 [67]. While previous paradigms improved efficiency and customization, Industry 6.0 demands agility to support advanced customization and sustainable resource management. This

shift involves adapting processes to enable mass customization and sustainable production models [68]. Using advanced technologies to collect detailed data on consumer preferences, organizations can deliver tailored products while maintaining environmental responsibility [69].

As a result, decentralized and dynamic structures empower organizations to meet market needs more rapidly, promoting continuous innovation and efficient supply chain management, which surpasses the capabilities of earlier industrial paradigms [70]. Based on these findings and comparisons, the following proposition can be stated:

- P9: *Compared to the more hierarchical structures of Industry 4.0 and the limited flexibility of Industry 5.0, the decentralized organizational structures of Industry 6.0 increase the agility of companies, enabling them to respond more quickly and flexibly to market fluctuations and local operating conditions through autonomous decision-making [61,62].*
- P10: *Structural adaptation to market dynamics and specific consumer preferences through Industry 6.0 enables organizations to implement mass customization models and sustainable resource management, exceeding the levels of customization and adaptability achievable in Industry 4.0 and Industry 5.0 [67,69].*

The Haier RenDanHeYi model (RenDanHeYi management model, Available online: https://www.rendanheyi.com/; accessed on 29 August 2025) illustrates decentralized cognitive architecture, where independent business units use AI and machine learning to analyze market data and make strategic decisions. Each unit acts as a "cognitive enterprise," autonomously handling product customization, supplier selection, and sustainability, showing how decentralization enhances competitiveness through cognitive capabilities.

3.4. Strategic Shifts

Industry 6.0 demands significant strategic adjustments, placing mass customization and sustainable long-term planning at the center of business models [71]. This shift requires organizations to redefine their production and market strategies, not only to meet evolving consumer expectations but also to establish sustainability as a core strategic objective [72]. Mass customization is a crucial element in Industry 6.0, as it compels companies to revise both their manufacturing processes and marketing approaches [73]. As consumer preferences increasingly focus on personalized solutions, organizations are required to serve individual needs while preserving operational efficiency [37]. Achieving this level of customization involves strategies that balance flexibility with scalability, allowing companies to adapt rapidly to changing consumer demands [21]. Moreover, marketing must also evolve to communicate individualized value effectively, positioning customization as a key differentiator and enhancing customer experience, which is now central to competitiveness [74].

Another strategic priority in Industry 6.0 is long-term sustainable planning, which extends the concept of sustainability beyond emissions reduction to encompass the entire product lifecycle and promote circular economy practices [45]. This approach emphasizes responsible resource management and considers both the environmental and social impacts of production activities [75]. The integration of AI and IIoT enables real-time monitoring and optimization of processes, which allows companies to minimize waste and continually improve resource efficiency. Sustainability, therefore, becomes a proactive responsibility throughout the value chain, encouraging collaboration with suppliers and partners who are also committed to sustainable practices [76]. As a result, Industry 6.0 supports investments in technologies and business processes that ensure responsible resource management and foster a sustainability-oriented corporate culture. This orientation, in turn, enhances resilience and long-term competitiveness [77].

Ultimately, Industry 6.0 guides organizations toward strategies that integrate personalization with sustainability. This dual focus transforms production and marketing models to align with the requirements of a more resilient and sustainable economy. Such a foundation enables modern organizations to thrive in increasingly dynamic markets. Based on these findings, we can state:

- P11: *Mass customization, enabled by the integration of AI and IIoT in Industry 6.0, requires a business strategy that balances operational flexibility and scalability, enhancing the ability of companies to respond to dynamic consumer preferences [21,69].*
- P12: *Integrating sustainability as a guiding principle of business strategy in Industry 6.0 fosters long-term value creation by promoting responsible resource management practices and a sustainability-oriented supply chain [45,75,76].*

Nike's cognitive manufacturing platform (Nike Willy Wonka factory: https://about.nike.com/en/stories/nike-spaces-sports-design-innovation-performance) leverages AI to optimize both product personalization and sustainability. By processing customer data alongside environmental factors like material use and emissions, the system autonomously balances customization with eco-friendly practices, illustrating how cognitive systems can harmonize individual preferences and environmental goals.

3.5. Challenges and Opportunities

Industry 6.0 is fundamentally oriented towards achieving a zero-defect manufacturing environment. This paradigm shift presents significant opportunities but also introduces notable challenges. A primary challenge lies in the cultural transformation required within organizations [39]. Unlike previous industrial paradigms where quality control was often reactive, Industry 6.0 emphasizes proactive, data-driven quality assurance. Such an approach fosters continuous improvement and supports incremental innovation [56]. Autonomous systems and real-time data analytics play a central role by enabling predictive quality management. These technologies allow companies to prevent defects before they occur, rather than simply correcting them after production. This shift leads to reduced waste and lower operational costs, while making production processes more resilient. Potential issues are identified and addressed early, minimizing disruption [78]. To fully realize these benefits, organizations must embrace a culture that prioritizes ongoing quality enhancement. Teams should be empowered to make informed decisions using real-time data [79]. This empowerment is essential for the agility and adaptability that Industry 6.0 requires, enabling decision-making at the operational level. Furthermore, the establishment of a zero-defect environment highlights the transition from static to dynamic quality management. In this context, quality is seen not just as a matter of compliance, but as a strategic objective that continuously drives innovation and adaptability [37]. Organizations are encouraged to cultivate a culture of incremental improvement, leveraging data to refine processes and respond swiftly to evolving market needs [1]. This cultural evolution enhances operational agility and strengthens customer trust and brand reputation, aligning with the overarching aims of Industry 6.0, particularly its commitment to sustainability. Recent contributions also highlight how AI and blockchain are being applied to quality management [16] and how neural network models can predict equipment faults in real time [17], reinforcing the role of predictive analytics in achieving zero-defect environments. Drawing upon these findings and comparisons, it is possible to affirm that:

- P13: *Implementing a zero-defect manufacturing environment in Industry 6.0 requires significant cultural changes that promote a culture of continuous quality and incremental improvement that supports continuous innovation within industrial organizations [24,79].*

- P14: *Establishing a zero-defect environment in Industry 6.0 strategically impacts organizations by promoting data-driven decision-making and proactive quality management, thereby increasing organizational agility and resilience [19,20].*

As an illustrative example of cognitive zero-defect cultural transformation, Toyota's advanced quality systems (Toyota Production System, Available online: https://global.toyota/en/company/vision-and-philosophy/production-system/; accessed on 29 August 2025) highlight how AI-driven predictive quality management can reshape an organization's approach from reactive problem solving to proactive quality assurance. Employees collaborate with cognitive systems that continuously analyze quality patterns, recommend process enhancements, and autonomously implement minor adjustments. This integration fosters a shift from traditional human-centered continuous improvement (kaizen) to a collaborative dynamic between human insight and machine intelligence, supporting the pursuit of zero-defect outcomes.

To enhance transparency and strengthen the theoretical grounding of our framework, Appendix A provides a detailed mapping between each proposition (P1–P14) and the corresponding references, clarifying how prior studies informed and supported the formulation of each proposition.

4. Towards Industry 6.0: Conceptual Models

This section presents theoretical models to clarify the transition from Industry 4.0 to Industry 6.0, highlighting major phases and changes. Building on earlier propositions (P1–P6), these models structure the analysis of technological, organizational, strategic, and cultural evolution. An evolutionary model outlines the main stages, while a conceptual model for Industry 6.0 examines key factors such as advanced technology, decentralized organization, mass customization, sustainability, cultural shifts, and innovation. The section concludes with practical guidelines for organizations moving toward Industry 6.0, aiming for clarity and analytical rigor.

4.1. Evolutionary Model from Industry 4.0 to Industry 6.0

The evolutionary model from Industry 4.0 to 6.0 depicts industrial transformation as a network of connected paradigms, each with five main elements: technology, organizational structure, business strategy, culture, and innovation (Figure 1). It should be noted that elements such as AI, IIoT, sustainability, and mass customization have appeared in earlier industrial paradigms, particularly Industry 4.0 and 5.0. However, in Industry 6.0, these components are not treated as isolated or incremental extensions but are reconfigured into a cyclical, self-reinforcing system. The novelty of the model lies in the feedback loops and in the cultural transformation that binds these elements together, making them operate as interdependent drivers of resilience and adaptability. These components shape every phase, showing that progress is both linear and qualitative, with interdependent relationships. Transition is driven primarily by technological advances. For instance, Industry 4.0 introduced IoT and Cyber–Physical Systems, enabling automation and digitalization to improve efficiency and cut costs.

As Industry 5.0 emerges, human–machine integration deepens with collaborative robotics and AI enabling more adaptable and customized production. This phase is marked by agile manufacturing that meets specific customer needs. Industry 6.0 brings fully autonomous, intelligent networks powered by advanced AI and IIoT, closely linking technology, sustainability, and organizational agility for higher efficiency.

Organizational structures evolve from centralized hierarchies in Industry 4.0 to greater internal collaboration in 5.0, facilitated by improved technological communication. In 6.0, organizations adopt decentralized structures that enable rapid decision-making at

operational levels through real-time data access. The model reflects this with interconnected influences between technology and organizational components, underlining the importance of decentralization.

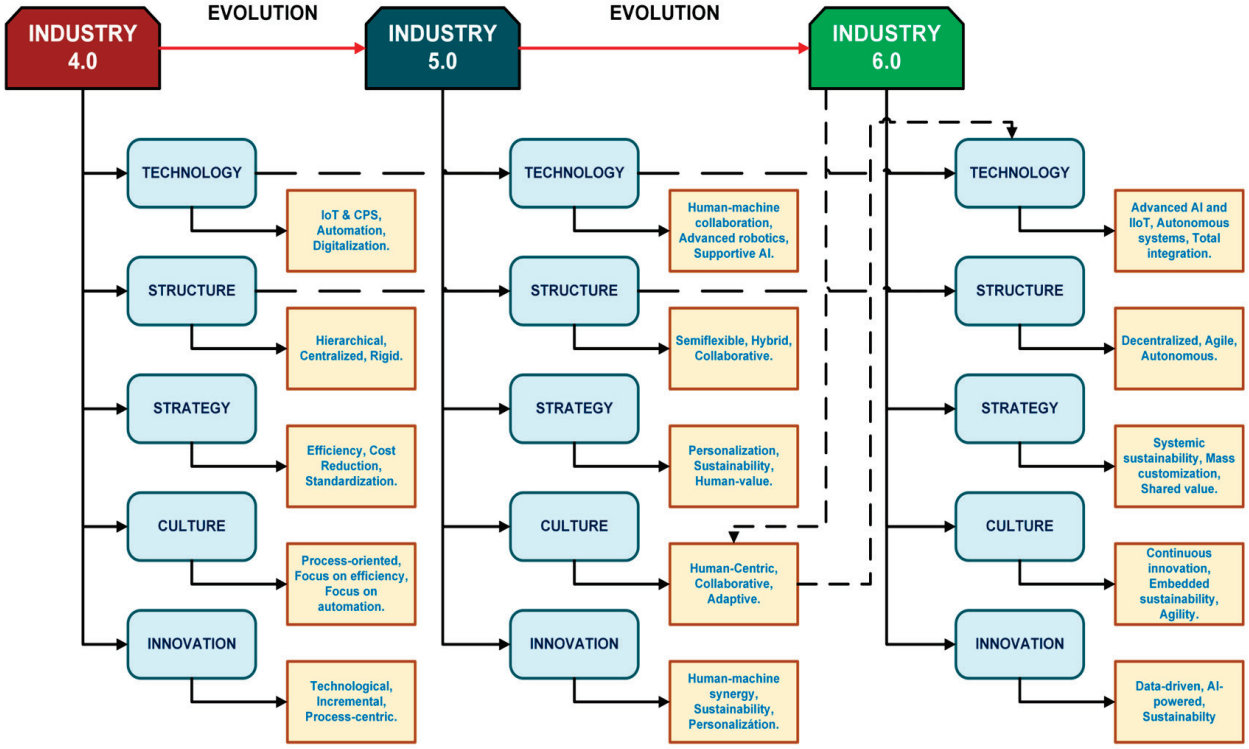


Figure 1. Graphical model for moving from Industry 4.0 to Industry 6.0 (Source: author's elaboration).

Strategically, Industry 4.0 focuses on efficiency and standardization, while 5.0 shifts toward customization and innovation, though sustainability remains secondary. By Industry 6.0, mass customization and sustainability take center stage, with strategies geared toward tailored solutions and responsible practices. Organizational culture transitions from process optimization to continuous improvement and proactive innovation, supporting adaptability and resilience in the face of constant technological change.

To summarize the main changes in technology, structure, strategy, culture, and innovation across the three industrial paradigms, Table 2 presents six Knowledge Insights that illustrate the evolution from Industry 4.0 to Industry 6.0, corresponding to propositions P1 to P6 discussed earlier in Section 3.1.

Table 2. Knowledge Insights on the evolution from Industry 4.0 to Industry 6.0.

Proposition	Knowledge Insights	Explanatory Notes
P1	Industry 4.0 relies on cyber–physical systems (CPS) and IoT for digital automation but maintains centralized structures.	This insight emphasizes the technological advancements of Industry 4.0, focusing on automation through CPS and IoT, but highlights that organizational structures remain centralized, limiting flexibility and agility.
P2	Industry 5.0 emphasizes human–machine collaboration and early sustainability integration, with minimal structural changes.	This insight underscores the shift towards human-centered technologies in Industry 5.0, integrating collaborative robots and AI, while suggesting that the organizational structure remains largely intact.

Table 2. *Cont.*

Proposition	Knowledge Insights	Explanatory Notes
P3	Industry 6.0 uses AI and IIoT to enable autonomous systems, requiring decentralized structures for agility and responsiveness.	The key difference in Industry 6.0 is the integration of fully autonomous systems, where decentralized decision-making structures are critical to ensure responsiveness and adaptability to rapid market changes.
P4	In Industry 6.0, business strategies prioritize mass customization and sustainability, overcoming prior efficiency-focused models.	This insight highlights how Industry 6.0 overcomes the limits of traditional efficiency-driven models, focusing on sustainable, tailored products that meet customer-specific demands while reducing environmental impact.
P5	A cultural shift is essential in Industry 6.0, fostering continuous improvement and integrating sustainability across the organization.	The cultural transformation in Industry 6.0 is critical to maintaining an innovation-driven environment. Sustainability becomes ingrained not just as a product feature, but as a core organizational value.
P6	Innovation in Industry 6.0 is proactive and data-driven, using advanced analytics and autonomous systems for continuous improvement.	Innovation in Industry 6.0 is rooted in advanced analytics and AI-driven systems, enabling real-time data collection and predictive analytics to drive continuous improvement in processes, products, and business strategies.

4.2. Conceptual Model from Industry 4.0 to Industry 6.0

The conceptual model of industrial evolution from Industry 4.0 to 6.0 is based on six propositions and six knowledge insights (see Table 1), highlighting key changes shown in Figure 2. It covers not just technology, but also the development of organizations, strategies, and cultures. P1 explains that cyber–physical systems (CPS) and IoT-enabled digital automation in Industry 4.0 without shifting away from centralized structures or efficiency-driven strategies. P2 notes that Industry 5.0 has increased human–machine collaboration, customization, and sustainability efforts, yet centralization remains dominant and sustainability is often seen as an add-on rather than a core strategy.

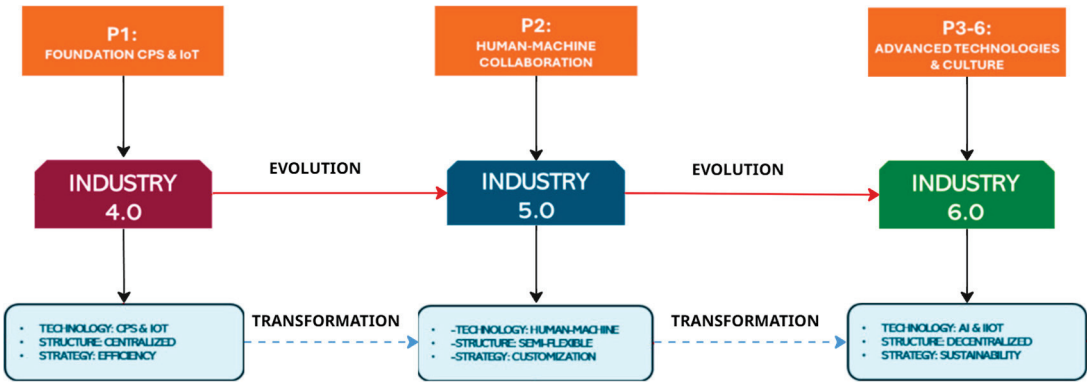


Figure 2. Conceptual model for moving from Industry 4.0 to Industry 6.0 (Source: author's elaboration).

The propositions P3 to P6 outline the distinctive features of Industry 6.0 compared to earlier paradigms. P3 highlights the central role of artificial intelligence and IIoT, which enable autonomous systems and require decentralized structures, thus increasing agility and responsiveness. P4 focuses on business strategies integrating mass customization with sustainability. P5 emphasizes the importance of ongoing cultural transformation, with

sustainability established as a core value. Finally, P6 shows that innovation in Industry 6.0 is proactive and data-driven, fostering technological development and the ability to anticipate market trends. The conceptual model illustrates a network in which changes in any component affect the entire system, demonstrating the need for a systemic and integrated approach to fully unlock the benefits of Industry 6.0.

4.3. Conceptual Model of Industry 6.0

The conceptual model of Industry 6.0, as illustrated in Figure 3, presents an interconnected and dynamic structure based on Propositions P7 to P14 detailed in Sections 3.2–3.5. This model integrates Artificial Intelligence (AI) and the Industrial Internet of Things (IIoT), which together drive significant transformations as highlighted in Propositions P7 and P8. Relationships within the model are depicted using solid and dashed arrows. Solid arrows indicate direct links and main propositions, whereas dashed arrows represent feedback loops and indirect influences among components.

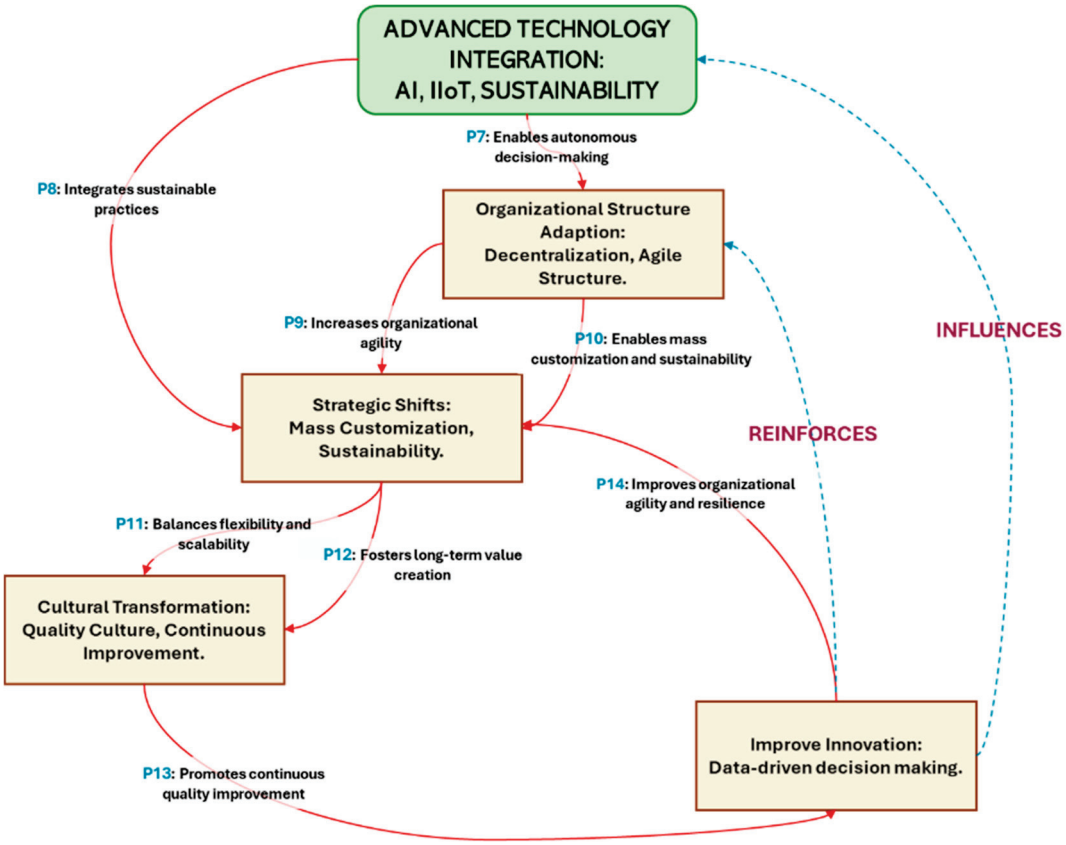


Figure 3. Conceptual model of Industry 6.0 (Source: author's elaboration).

At the core of Industry 6.0 lies Advanced Technology Integration. Here, AI and IIoT form the foundation for systems capable of autonomous decision-making and predictive optimization. Proposition P7 demonstrates how these technologies support organizational adaptation and enable intelligent automation for data-driven decision-making. AI facilitates continuous machine learning and adaptation, while IIoT connects devices and systems, enabling real-time communication and information exchange. Proposition P8 explains how technological integration supports strategic renewal, combining sustainable practices with the capacity for mass customization. Advanced technologies facilitate the collection and analysis of extensive consumer and operational data. This supports strategies that merge personalized product offerings with operational efficiency and environmental stewardship. As a result of technological advancement, organizations increasingly adopt decentralized

and agile structures. Once these structures are implemented, companies gain greater flexibility and responsiveness, as described in Propositions P9 and P10. Proposition P9 clarifies that decentralization enables faster adaptation to market changes, allowing business units to make decisions based on localized data. This capability improves organizational agility and responsiveness to shifts in consumer preferences or market conditions. Proposition P10 highlights that flexible organizational design supports the delivery of customized products while managing resources sustainably. Decentralization also encourages collaboration across departments and organizational levels, fostering innovation and resource efficiency. This approach ensures that specific customer needs are met without compromising the overall sustainability objectives of the enterprise.

The Industry 6.0 model is grounded in strategic change facilitated by advanced technologies and decentralized organizational structures. Propositions P11 and P12 outline how organizations develop strategies that enhance adaptability to consumer demands. Proposition P11 examines how business strategy shapes the balance between customization and efficiency, emphasizing the importance of quickly adapting products and services to customer requirements while maintaining high levels of productivity. Proposition P12 addresses the integration of sustainability into core business practices, supporting the creation of long-term value. Organizations now adopt sustainable approaches not only to comply with environmental standards but also to respond to increasing consumer expectations for social and environmental responsibility. This strategic orientation enhances a firm's competitive position and strengthens its reputation.

Cultural transformation is a critical component of the model, focusing on ongoing quality improvement. Proposition P13 demonstrates how attention to quality encourages innovation and fosters a culture of continuous advancement. Such a culture motivates personnel to develop new ideas, experiment with novel solutions, and collaborate across organizational boundaries. Proposition P14 details how strategic modifications can enhance a company's responsiveness to market changes, technological developments, and evolving customer needs, thereby ensuring sustained competitiveness. The Industry 6.0 conceptual model highlights a complex network of interrelated and circular connections among its components. Solid arrows represent direct and immediate relationships, while dashed arrows indicate feedback loops and longer-term influences.

1. Feedback from Innovation to Organizational Structure: Innovation is not only influenced by organizational structure but in turn can stimulate structural changes. For example, new ideas or technologies may require internal reorganization to be implemented effectively.
2. Feedback from Innovation to Technology Integration: Innovation can lead to the development or adoption of new technologies, further fueling the cycle of technological advancement.

These interconnections ensure that the system remains in a state of ongoing transformation, fostering an adaptive and resilient organizational ecosystem. Each element of the model both influences and is influenced by the others, highlighting the importance of approaching Industry 6.0 as an integrated system. The distinction between continuous and dashed arrows clarifies the nature of the transformations: direct relationships represent immediate and intentional impacts, while feedback loops reflect longer-term effects resulting from ongoing interactions among system components. The model's strength lies in its circular and synergistic structure. Digital and sustainable transformation arises not from isolated changes, but from the dynamic interplay among all components. This integrated perspective is essential for fully leveraging the value of Industry 6.0, as it enables organizations to innovate, remain agile, and pursue sustainability in a highly competitive environment.

5. Building Theory for Industry 6.0

The development of a theoretical framework for Industry 6.0 is essential for understanding and guiding the transition to this new industrial paradigm. In this section, we apply a theory-building methodology [80] to construct a comprehensive theoretical model that synthesizes the main components and their interconnections, as identified in propositions P1–P14 from the previous sections. Theory building is a systematic process for formulating a coherent and robust theory by defining key constructs, establishing relationships among them, and integrating these elements into a unified conceptual model [81]. This approach enables the logical organization of complex ideas, making it easier to understand the relevant phenomena and serving as a foundation for future research and practical applications [82]. A construct is a key theoretical concept central to the study's focus. In Industry 6.0, constructs stem from propositions P1–P14, reflecting major technological, organizational, strategic, and cultural changes. These include technology integration, decentralized structures, personalization and sustainability strategies, cultural shifts, and innovation enhancement. Identifying these constructs requires analyzing and synthesizing the main ideas from each proposition.

- C1. Advanced Technology Integration emerges from the propositions that highlight the importance of technologies such as Artificial Intelligence (AI) and Industrial Internet of Things (IIoT) in enabling autonomous systems and improving operational efficiency (P3, P7, P8).
- C2. Decentralized Organizational Structures are identified as necessary to take full advantage of new technologies and increase organizational agility, as discussed in propositions P3, P9, and P10.
- C3. Mass Personalization and Sustainability Strategies are derived from propositions that emphasize the importance of balancing operational flexibility, personalization, and environmental responsibility (P4, P11, P12).
- C4. Cultural Transformation is a construct that emerges from the need to foster a culture of continuous improvement and integrate sustainability values, as described in P5 and P13.
- C5. Boosting Innovation is identified as a key element to drive continuous improvements in technology strategy and adoption, based on propositions P6 and P14.

In the process of developing a theoretical framework, once the constructs have been defined, we proceed as follows:

1. Elaboration of Relationships: It is necessary to establish how these constructs interact with one another, identifying causal or mutually influential relationships.
2. Formulation of the Theoretical Model: The next step involves integrating the constructs and relationships into a coherent structure that represents the phenomenon under study.
3. Theoretical Validation: The model should then be validated by linking it to existing literature. This entails discussing how the model contributes to the field of study and what implications it holds.
4. Implications and Future Perspectives: It is important to explore the practical consequences of the model and identify potential avenues for further research.

This approach establishes a strong theoretical framework for Industry 6.0, clarifying changes in industry and offering guidance to organizations shifting to the new paradigm. It explains how advanced technologies, structures, strategies, culture, and innovation shape adaptive, sustainable systems. The next section defines main concepts, outlines their relationships, and introduces an integrated Industry 6.0 model, relating it to literature and emphasizing its practical value.

5.1. Processing Construct Relationships

The identified constructs represent the fundamental elements of Industry 6.0. To gain a comprehensive understanding of the way these elements interact and influence industrial transformation, it is imperative to establish the causal relationships and mutual influences that exist between them.

- C1. Advanced Technology Integration → C2. Decentralized Organizational Structures. Advanced Technological Integration (C1) functions as a catalyst for the transformation of organizational structures. The implementation of technologies such as AI and the Industrial Internet of Things (IIoT) enables autonomous systems and data-driven decision-making in real time. This capability necessitates and encourages the development of decentralized organizational structures (C2), as evidenced by propositions P3, P7, and P9. The implementation of advanced technologies enables operational departments to make autonomous decisions, thereby reducing the necessity for centralized control and increasing organizational agility.
- C2. Decentralized Organizational Structures → C3. Mass Personalization and Sustainability Strategies. The implementation of Decentralized Organizational Structures (C2) directly influences Mass Personalization and Sustainability Strategies (C3). As previously discussed in Propositions P4, P10, P11, and P12, the decentralization of organizational structures allows for greater operational flexibility, enabling the organization to quickly adapt to dynamic consumer preferences and implement sustainable practices. Autonomous business units can tailor products and services to local needs, thereby balancing efficiency and environmental responsibility.
- C3. Mass Personalization Strategies and Sustainability → C4. Cultural Transformation. The implementation of Mass Personalization and Sustainability Strategies (C3) necessitates a comprehensive cultural transformation (C4). Propositions P5, P12, and P13 illustrate the necessity of an organizational culture that fosters continuous improvement, innovation, and the adoption of sustainable values to effectively implement strategies focused on personalization and sustainability. The corporate culture must transform to align with these novel strategic objectives, necessitating a culture that fosters continuous learning and adaptability.
- C4. Cultural Transformation → C5. Boosting Innovation. Cultural transformation (C4) is a prerequisite for the empowerment of innovation (C5). An innovation-oriented culture, as delineated in Propositions P6 and P14, cultivates the germination of novel concepts and the incorporation of nascent technologies. The promotion of values such as creativity, collaboration, and openness to change within the organization stimulates proactive, data-driven innovation, which is essential for maintaining a competitive advantage in Industry 6.0.
- C5. Enhancing Innovation → C1. Advanced Technology Integration and C2. Decentralized Organizational Structures. The enhancement of innovation (C5) generates a positive feedback loop with advanced technological integration (C1) and decentralized organizational structures (C2). The continued development and adoption of new technologies is a consequence of sustained innovation, which in turn drives further integration of technology (P6). Moreover, it may necessitate modifications to organizational structures to facilitate the implementation of novel innovative processes and systems (P14), thereby establishing a virtuous cycle of continuous improvement and adaptation.

5.2. Building the Integrated Theoretical Model

By integrating the identified constructs and their relationships, the Integrated Theoretical Model of Industry 6.0 is outlined (Figure 4). This model illustrates a dynamic

and interconnected system where each component influences the others, generating a continuous cycle of transformation and improvement. The model shows that successfully adopting Industry 6.0 requires a holistic approach. In this perspective, technology, organizational structure, business strategy, culture, and innovation are all interdependent elements that must be aligned for organizations to thrive in the new industrial context. The transformation process begins with the integration of advanced technologies (C1), which forms the foundation of Industry 6.0. The introduction of tools such as artificial intelligence (AI) and the industrial Internet of Things (IIoT) enables the creation of autonomous systems and the optimization of processes based on prediction. These technologies support intelligent automation, allowing machines to execute complex tasks independently of human intervention. They can adapt and learn from experience by using sophisticated machine learning algorithms. Furthermore, these systems facilitate real-time data analytics, as the IIoT makes it possible to collect and process significant amounts of operational data, improving visibility and control. Advanced analytics also support data-driven decisions, with AI helping organizations reduce errors and increase overall efficiency. This phase is crucial because it provides the technical foundation needed for further transformation. For organizations to fully capitalize on the potential of Industry 6.0, they must ensure the effective integration of advanced technologies. The adoption of these technologies also makes it necessary to adapt organizational structures accordingly.

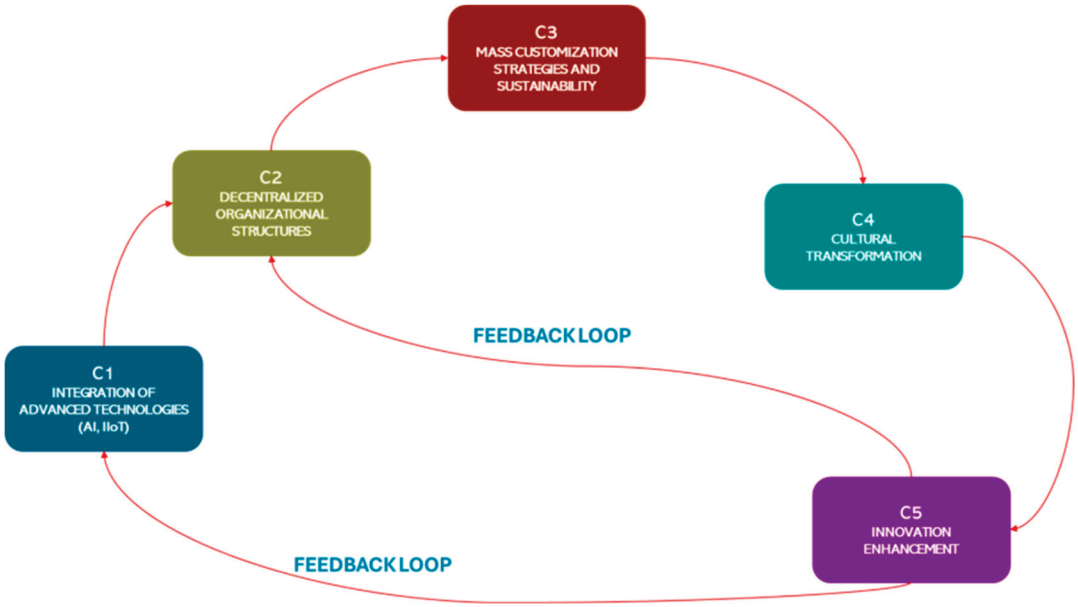


Figure 4. Industry 6.0 theoretical model (Source: author's elaboration).

The emergence of decentralized organizational structures (C2) directly results from the need to leverage operational autonomy. These structures allow departments and teams to make rapid, informed decisions based on local data, thereby increasing organizational agility. By reducing hierarchical levels, companies can respond more swiftly to market changes. Furthermore, the flatter organizational setup encourages collaboration, facilitates information sharing, and supports cross-functional innovation. The decentralization process is enabled by technologies introduced in the initial stage, which provide the tools necessary for effective communication and autonomous process management.

Establishing appropriate organizational structures makes it possible to implement Mass Personalization and Sustainability Strategies (C3). This involves using data-driven and technology-enabled mass customization, offering personalized products and services on a large scale without increasing costs. At the same time, sustainability is integrated into all operational aspects, from resource management to product design. These strategies

address the evolving expectations of consumers, who increasingly desire personalized offerings and demonstrate heightened environmental awareness. Adopting such strategies helps companies differentiate themselves in the market and generate long-term value.

To implement these strategies successfully, it is essential to transform organizational routines and mindsets. Employees should be encouraged to seek continuous improvement in processes and products, while the company's mission and values must reflect sustainability and a commitment to innovation. Company culture serves as the binding element that supports both technological advancements and changes in work practices. When organizational culture is transformed, it leads to further enhancement of innovation (C5). At this stage, the company not only responds to change but also anticipates trends and drives innovation within its sector. A data-driven approach enables the organization to identify new opportunities and optimize operations. Developing innovative solutions empowers the company to create advanced products, services, and business models. This, in turn, provides a competitive advantage and sustains the cycle of transformation. The model incorporates a feedback loop that connects innovation enhancement (C5) with advanced technology integration (C1) and decentralized organizational structures (C2). This loop implies that the introduction of new technologies stimulates further adoption and supports business operations; changes in organizational structure align with the introduction of innovative technologies and strategies, potentially requiring additional structural adjustments; organizational culture evolves to support new initiatives while maintaining a focus on ongoing improvement and innovation. This continuous process ensures that the company remains adaptable and resilient, capable of responding to market challenges and opportunities. The graphical representation of the Integrated Theoretical Model, shown through a circular or spiral diagram (Figure 4), illustrates the cyclical and progressive nature of this process. The constructs (C1–C5) are presented sequentially, with direct relationships indicated by arrows.

- C1 is a prerequisite for C2. The implementation of advanced technological integration necessitates the establishment of decentralized organizational structures.
- C2 $\rightarrow$ C3: Decentralized structures facilitate the adoption of mass personalization and sustainability strategies.
- C3 $\rightarrow$ C4: The implementation of these strategies facilitates cultural transformation.
- C4 $\rightarrow$ C5: Cultural transformation drives the empowerment of innovation.
- C5 $\rightarrow$ C1/C2: Enhancing innovation generates new technological developments and necessitates further organizational adaptations, thereby closing the feedback loop.

The feedback arrows underscore the influence of each stage on the others, creating an interactive and self-reinforcing system.

5.3. Theoretical Validation of the Model

The integrated theoretical model of Industry 6.0 is grounded in several established management theories, connecting traditional concepts with the requirements of contemporary industrial change. According to Contingency Theory [83], organizations should adapt their structures to match technological and environmental conditions. Within this model, Advanced Technological Integration (C1) calls for Decentralized Organizational Structures (C2) to fully realize the potential of new technologies, reflecting the contingency approach. The Resource-Based View [84] highlights that organizations gain a competitive edge by cultivating resources that are rare and difficult to imitate. In our framework, Advanced Technology Implementation (C1) and Innovation Enhancement (C5) serve as such strategic resources. Dynamic Capabilities Theory [85] stresses the importance of developing the ability to respond quickly to changes. Our model shows that decentralized structures (C2) and a focus on cultural transformation (C4) are central to building these

capabilities. Principles from Organizational Learning Theory [86] are visible in the Cultural Transformation (C4) and Innovation Enhancement (C5) components, which together foster a culture of ongoing learning and innovation. Systems Theory [87] views organizations as interconnected entities. The model illustrates this by showing how technology (C1), organizational structure (C2), strategy (C3), culture (C4), and innovation (C5) interact and influence one another. Transformational Leadership Theory [88] emphasizes the role of leaders in driving organizational change. In the context of Industry 6.0, effective leadership is crucial to ensuring successful cultural transformation (C4). Stakeholder Theory [89] advocates for considering the interests of all stakeholders. The model addresses this through mass personalization and sustainability strategies (C3), which account for the needs of customers, communities, and the environment. Porter's Competitive Advantage Theory [90] is reflected in the use of advanced technologies (C1) and innovation (C5) to differentiate and optimize organizational performance. The model incorporates these elements through constructs C1, C2, and C3. The framework's complexity arises from its dynamic and adaptive nature, enabling organizations to evolve in response to change through feedback loops that connect the various constructs. Knowledge Management Theory [91] is also present in the Innovation Enhancement (C5) and Cultural Transformation (C4) components, demonstrating the critical role of knowledge management in promoting continuous innovation. Finally, our model not only aligns with these foundational theories but also integrates them to support the effective adoption of Industry 6.0. It demonstrates how technology, structure, strategy, culture, and innovation are interrelated and collectively enhance organizational adaptability.

5.4. Final Theoretical Contribution

The ultimate theoretical contribution of this study is the formulation of a central theoretical proposition that encapsulates the core tenets of the integrated model of Industry 6.0. This proposition emphasizes the interdependence of constructs and the cyclical nature of transformation. The final theoretical assertion is as follows:

- *Success in Industry 6.0 is not achieved merely through the adoption of advanced technologies or the implementation of new strategies but emerges from a cyclical and interdependent transformation wherein Advanced Technological Integration (C1) reconfigures Organizational Structures (C2), which in turn enable and are reshaped by Mass Personalization and Sustainability Strategies (C3). This dynamic process initiates a profound Cultural Transformation (C4) that elevates Innovation (C5) from an operational function to an intrinsic strategic capability. This enhanced innovation further fuels technological integration and organizational evolution, establishing a virtuous cycle that redefines the organization's identity and operational essence, making agility, resilience, and sustainability not just objectives but fundamental components of its DNA.*

This proposition encapsulates the essence of the model, emphasizing that success in Industry 6.0 is contingent upon the synergistic effect of all identified components, rather than the result of isolated changes. Each construct is intimately connected and mutually reinforcing:

Advanced Technology Integration (C1) serves as the initial catalyst, enabling new operational capabilities through technologies like AI and IIoT. However, to fully realize its potential, it requires Decentralized Organizational Structures (C2) that allow for agility and rapid decision-making at all levels. Decentralized Organizational Structures (C2) facilitate the implementation of Mass Personalization and Sustainability Strategies (C3), empowering autonomous units to tailor products and services to local needs while balancing efficiency with environmental responsibility. The effectiveness of these strategies (C3) hinges on a Cultural Transformation (C4) that fosters a culture of continuous improvement, innovation,

and the integration of sustainability values throughout the organization. This Cultural Transformation (C4) propels Innovation Enhancement (C5), stimulating the generation of new ideas and the adoption of emerging technologies, which are essential for maintaining a competitive advantage. Innovation Enhancement (C5) creates a positive feedback loop by driving further Advanced Technology Integration (C1) and necessitating additional adaptations in Decentralized Organizational Structures (C2), thereby perpetuating the cycle of continuous improvement and adaptation.

This cyclical and interdependent model underscores that success in Industry 6.0 requires a holistic approach where technological advancements, organizational restructuring, strategic initiatives, cultural shifts, and innovation practices are aligned and mutually reinforcing. By embracing this integrated framework, organizations can navigate the complexities of the modern industrial landscape, fostering agility, resilience, and sustainability [92] essential for long-term success.

This proposition captures the essence of the model, highlighting that the success of Industry 6.0 depends on the synergistic interaction of all its components rather than isolated changes. Each construct is closely linked and reinforces the others. Advanced Technology Integration (C1) acts as the primary catalyst, providing new operational capabilities through technologies such as AI and IIoT. However, its full potential is realized only when supported by Decentralized Organizational Structures (C2), which enable agility and timely decision-making throughout the organization. These decentralized structures also facilitate the adoption of Mass Personalization and Sustainability Strategies (C3), empowering autonomous units to respond to local needs while balancing efficiency and environmental stewardship.

The effectiveness of these strategies (C3) depends on a genuine Cultural Transformation (C4), which nurtures continuous improvement, innovation, and the integration of sustainability values at every level. This cultural evolution (C4) drives Innovation Enhancement (C5), supporting the development of new ideas and the implementation of emerging technologies, which are essential for maintaining competitiveness. Innovation Enhancement (C5) then reinforces the cycle by promoting further Advanced Technology Integration (C1) and prompting new adaptations in Decentralized Organizational Structures (C2), ensuring ongoing improvement and organizational flexibility. This cyclical and interdependent approach demonstrates that Industry 6.0 requires a comprehensive framework, where advancements in technology, organizational changes, strategic initiatives, cultural development, and innovation practices are aligned and mutually supportive. By adopting this integrated perspective, organizations can effectively address the challenges of the contemporary industrial environment, strengthening agility, resilience, and sustainability as core elements for long-term success.

To this end, our proposed framework distinguishes itself from previous approaches by moving beyond a focus on standalone technological enablers, such as multi-agent [7] or digital twin integration [8], and instead embeds these elements within a systemic, cyclical perspective in which feedback loops and cultural transformation play a central role.

6. Discussion of Results

This section discusses the theoretical study's results in relation to the research questions. The integrated Industry 6.0 model developed here highlights essential technological, organizational, strategic, and cultural changes for this emerging industrial paradigm. To ground the theoretical contribution in emerging industrial practice, we included several illustrative cases throughout the analysis. For instance, Siemens' Amberg factory demonstrates cognitive quality control, where AI autonomously adjusts production parameters in real time. The Haier RenDanHeYi model exemplifies decentralized organizational architec-

tures that empower independent units with cognitive decision-making capabilities. Nike's cognitive manufacturing platform illustrates how AI can harmonize personalization and sustainability, while Toyota's predictive quality systems show how zero-defect principles can reshape organizational culture. These examples, while not exhaustive, support the plausibility of the proposed framework by showing how elements of Industry 6.0 are already materializing in advanced manufacturing contexts.

Response to RQ1: How does Industry 6.0 influence the structure and strategy of industrial organizations?

Our findings indicate that Industry 6.0 reconfigures both organizational architecture and business strategy. Advanced technologies such as AI and IIoT (C1) act as catalysts for decentralization (C2), enabling real-time decision-making and enhanced agility. Strategic orientation shifts toward balancing mass customization and sustainability (C3), ensuring that organizations can deliver tailored products without compromising environmental responsibility [35,36]. Importantly, the model emphasizes that cultural transformation (C4) underpins these changes, embedding adaptability and continuous improvement across the enterprise. This holistic transformation surpasses prior frameworks that treated these domains separately, underscoring our contribution of a systemic, cyclical design. In particular, cultural transformation is not addressed as a generic change process but as a structural binding construct that continuously connects technology, organization, strategy, and innovation, distinguishing Industry 6.0 from previous maturity-oriented models.

Response to RQ2: What organizational structures are best suited to support the decentralized decision-making and integration of advanced technologies envisaged by Industry 6.0?

The results highlight decentralized organizational structures (C2) as most effective for leveraging the potential of Industry 6.0. These structures empower teams with autonomy, supported by real-time information from AI-driven control systems (C1). In practice, distributed additive manufacturing offers a clear example: local production cells autonomously adjust printing parameters, reducing reliance on central supervision and enhancing responsiveness. Such emerging practices confirm that decentralization, when combined with advanced technology, generates resilience and operational agility beyond what centralized structures can provide [62]. Similar directions are noted in recent studies where multi-agent systems are adopted to coordinate distributed manufacturing tasks [7] and digital product passport frameworks are integrated with digital twins to enable decentralized control [8].

Distributed additive manufacturing networks demonstrate cognitive decentralization by allowing local nodes to autonomously manage printing, material use, and quality control. These nodes adapt in real time to conditions like humidity and temperature, while inter-node communication ensures global quality and sustainability.

Response to RQ3: How can organizations develop strategies that effectively balance mass customization and sustainability within the Industry 6.0 model?

The model suggests that the balance between personalization and sustainability (C3) can be achieved by integrating advanced analytics with decentralized decision-making. AI and IIoT support mass customization while simultaneously optimizing resource use and minimizing waste [55]. This dual focus is increasingly evident in sectors such as predictive maintenance, where localized teams use machine learning insights to anticipate equipment failures and intervene proactively. By reducing downtime and avoiding unnecessary resource consumption, firms demonstrate how personalized service delivery and sustainable practices can reinforce each other within Industry 6.0 environments. This is consistent with neural-network-based approaches to fault prediction [17] and Six Sigma applications

enhanced by AI [16], which demonstrate how data-driven systems can embed resilience and sustainability into industrial operations.

Cognitive predictive maintenance systems balance customization and sustainability by using machine learning to schedule maintenance for each machine, optimizing resources facility-wide. These systems learn each machine's operation, predict the best times for intervention, and coordinate activities to reduce downtime and environmental impact.

Response to RQ4: What are the cultural and strategic implications of a zero-fault environment and how does it affect continuous innovation and incremental improvement in industrial organizations?

Achieving a zero-defect environment requires a cultural transformation (C4) that prioritizes quality, innovation, and preventive approaches over reactive corrections [21]. Strategically, this transformation strengthens innovation (C5), as predictive quality systems encourage experimentation and cross-functional collaboration. By embedding zero-defect objectives within organizational culture, firms move from static compliance to dynamic improvement, creating environments where incremental learning and radical innovation coexist. This cultural embedding distinguishes our model from prior accounts that largely treated zero-defect as a technological or operational goal rather than a cultural and strategic imperative.

Furthermore, this study addresses several gaps in existing literature:

1. This study investigates the impact of Industry 6.0 on organizational structures, with a focus on decentralization that enables distributed decision-making and the assimilation of new technologies.
2. In contrast to Industry 4.0 and 5.0, our framework positions sustainability as a core component, guiding organizations toward balancing mass customization with ecological responsibility.
3. We highlight necessary cultural transformations for achieving zero-defect environments, which are crucial for fostering sustained innovation and organizational resilience.
4. The theoretical model presented combines aspects of technology, organizational architecture, strategy, culture, and innovation, providing both actionable recommendations and fresh perspectives for management scholars and practitioners.

7. Conclusions

This study develops a conceptual framework for Industry 6.0, positioning it as an emerging paradigm that integrates advanced technologies, decentralized organizational forms, mass customization, sustainability, and innovation into a systemic and cyclical model. By highlighting the binding role of cultural transformation, the framework clarifies how technology, organization, strategy, and innovation are continuously connected and reinforced, moving beyond the linear or siloed representations of Industry 4.0 and 5.0.

The contribution of this research is threefold. First, it provides a clear working definition of Industry 6.0, addressing the current fragmentation in the literature. Second, it introduces propositions (P1–P14) that establish a theoretically grounded foundation for further inquiry. Third, it illustrates the plausibility of the framework through emerging industrial cases such as Siemens, Haier, Nike, and Toyota.

Several limitations must also be acknowledged. The applicability of the framework may be constrained in highly regulated industries, and SMEs may face barriers due to resource constraints and limited digital infrastructures. Moreover, the emphasis on cultural transformation presupposes organizational readiness that may vary across contexts, making diffusion uneven.

Overall, this research advances the debate on the future of manufacturing by framing Industry 6.0 as an adaptive and context-sensitive paradigm. Rather than replacing Industry

4.0 and 5.0, it provides an integrative lens for understanding how cognitive manufacturing, cultural transformation, and systemic feedback loops can drive more resilient, sustainable, and innovative industrial ecosystems. Thus, Industry 6.0 emerges not as a prescriptive blueprint but as a conceptual lens for rethinking how manufacturing systems can evolve toward greater resilience, sustainability, and cognitive adaptability.

Author Contributions: Conceptualization, A.F.-M. and D.S.-B.; methodology, S.O.-M. and M.J.-C.; investigation, A.F.-M. and M.J.-C.; data curation, A.P.F.d.H. and F.E.G.-M.; writing—original draft preparation, D.S.-B. and A.P.F.d.H.; validation and supervision, S.O.-M. and F.E.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original conceptual contributions presented in this study are included in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A Proposition–Reference Mapping

Proposition	Key Reference	Contribution of the Cited Works
P1	[32,37]	These works establish the foundational role of CPS and IoT in Industry 4.0, enabling digital automation but retaining centralized and efficiency-oriented structures.
P2	[34,36]	These studies on Industry 5.0 emphasize human–machine collaboration and sustainability awareness, but note limited changes in organizational structure, supporting the transition argument.
P3	[5,37]	Literature on AI and IIoT demonstrates how autonomous systems drive the need for decentralized organizations, showing a step beyond 4.0 and 5.0.
P4	[38,39]	These contributions highlight the strategic importance of mass customization and sustainability, positioning them as central objectives for Industry 6.0.
P5	[40,43]	Cultural change is discussed as a driver of innovation and sustainability in industrial transitions, but without systemic integration—our proposition builds on and extends these insights.
P6	[42]	This work stresses the shift toward proactive, data-driven innovation enabled by advanced analytics, aligning with our view of continuous innovation in Industry 6.0.
P7	[44,45]	These studies document how AI/IIoT integration improves efficiency and flexibility, forming the basis for Industry 6.0's predictive optimization and autonomous decision-making.
P8	[56,58]	Research on sustainable practices in advanced manufacturing shows how AI/IIoT can reconcile customization and environmental goals, supporting our proposition.
P9	[61,62]	Decentralization and agility are emphasized in these works, demonstrating how distributed structures enhance responsiveness compared to hierarchical models.
P10	[67,69]	These papers show how adaptation to market dynamics and consumer preferences requires new structural models, validating the Industry 6.0 approach.
P11	[21,69]	These contributions discuss how AI and IIoT enable scalable customization strategies, supporting the balance of flexibility and scalability proposed in P11.
P12	[45,75,76]	Sustainability as a strategic imperative is emphasized here, particularly the role of supply chains and responsible resource management in long-term value creation.
P13	[24,79]	These studies explore zero-defect manufacturing and cultural change for quality, underlining the organizational and cultural shifts required for implementation.
P14	[19,20]	These works provide evidence of predictive quality and defect prevention as strategic levers, showing their role in resilience and agility.2

References

1. Duggal, A.S.; Malik, P.K.; Gehlot, A.; Singh, R.; Gaba, G.S.; Masud, M.; Al-Amri, J.F. A Sequential Roadmap to Industry 6.0: Exploring Future Manufacturing Trends. *IET Commun.* **2022**, *16*, 521–531. [CrossRef]
2. Folgado, F.J.; Calderón, D.; González, I.; Calderón, A.J. Review of Industry 4.0 from the Perspective of Automation and Supervision Systems: Definitions, Architectures and Recent Trends. *Electronics* **2024**, *13*, 782. [CrossRef]
3. Ghobakhloo, M.; Iranmanesh, M.; Tseng, M.L.; Grybauskas, A.; Stefanini, A.; Amran, A. Behind the Definition of Industry 5.0: A Systematic Review of Technologies, Principles, Components, and Values. *J. Ind. Prod. Eng.* **2023**, *40*, 432–447. [CrossRef]
4. Fernández-Miguel, A.; García-Muiña, F.E.; Settembre-Blundo, D.; Tarantino, S.C.; Riccardi, M.P. Exploring Systemic Sustainability in Manufacturing: Geoanthropology's Strategic Lens Shaping Industry 6.0. *Glob. J. Flex. Syst. Manag.* **2024**, *25*, 579–600. [CrossRef]
5. Carayannis, E.G.; Kostis, P.C.; Kafka, K.I.; Valvi, T. Toward Industry 6.0 and Techno-Centric vs. Human-Centric Smart Balancing: The Journey from Industry and Society 4.0 toward Industry and Society 5.0 and Beyond. In *Sustainable Economic Development*; Routledge: Abingdon-on-Thames, UK, 2024; pp. 171–184. [CrossRef]
6. Enang, E.; Bashiri, M.; Jarvis, D. Exploring the Transition from Techno Centric Industry 4.0 towards Value Centric Industry 5.0: A Systematic Literature Review. *Int. J. Prod. Res.* **2023**, *61*, 7866–7902. [CrossRef]
7. Lou, S.; Hu, Z.; Zhang, Y.; Feng, Y.; Zhou, M.; Lv, C. Human-Cyber-Physical System for Industry 5.0: A Review From a Human-Centric Perspective. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 494–511. [CrossRef]
8. Bukowski, L.; Werbinska-Wojciechowska, S. Towards Maintenance 5.0: Resilience-Based Maintenance in AI-Driven Sustainable and Human-Centric Industrial Systems. *Sensors* **2025**, *25*, 5100. [CrossRef] [PubMed]
9. Rosário, A.T.; Raimundo, R. Internet of Things and Distributed Computing Systems in Business Models. *Future Internet* **2024**, *16*, 384. [CrossRef]
10. Srivastava, A.; Kumari, G.; Pathania, A. Technological Innovation and Accounting in Industry 6.0. In *Sustainable Innovation for Industry 6.0*; IGI Global: Hershey, PA, USA, 2024; pp. 286–309.
11. Sakurada, L.; De la Prieta, F.; Leitao, P. The Role of Multi-Agent Systems in Realizing Asset Administration Shell Type 3. *Future Internet* **2025**, *17*, 270. [CrossRef]
12. Melo, V.; de la Prieta, F.; Leitao, P. Relationship of Digital Product Passport and Digital Twin in Industry 4.0 Context. In *International Workshop on the Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*; Borangiu, T., Trentesaux, D., Leitão, P., Legat, C., Eds.; Springer: Cham, Switzerland, 2025; pp. 187–200.
13. Patrício, L.; Varela, L.; Silveira, Z.; Felgueiras, C.; Pereira, F. A Framework for Integrating Robotic Process Automation with Artificial Intelligence Applied to Industry 5.0. *Appl. Sci.* **2025**, *15*, 7402. [CrossRef]
14. Deshpande, S.; Jogdand, R. Development of IoT Middleware Broker Communication Architecture for Industrial Automation with Focus on Future Pandemic Possibilities: Industry 5.0. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022*; Springer Nature Singapore: Singapore, 2023; pp. 47–58. [CrossRef]
15. Islam, M.T.; Sepanloo, K.; Woo, S.; Woo, S.H.; Son, Y.-J. A Review of the Industry 4.0 to 5.0 Transition: Exploring the Intersection, Challenges, and Opportunities of Technology and Human–Machine Collaboration. *Machines* **2025**, *13*, 267. [CrossRef]
16. Li, L.; Duan, L. Human Centric Innovation at the Heart of Industry 5.0—Exploring Research Challenges and Opportunities. *Int. J. Prod. Res.* **2025**, 1–33. [CrossRef]
17. Fathi, M.; Ghobakhloo, M. Enabling Mass Customization and Manufacturing Sustainability in Industry 4.0 Context: A Novel Heuristic Algorithm for in-Plant Material Supply Optimization. *Sustainability* **2020**, *12*, 6669. [CrossRef]
18. Fedele, M.; Formisano, V. Waste from Criticality to Resource through an Innovative Circular Business Model: A Case Study in the Manufacturing Industry. *J. Clean. Prod.* **2023**, *407*, 137143. [CrossRef]
19. Najafi, B.; Najafi, A.; Farahmandian, A. The Impact of Artificial Intelligence and Blockchain on Six Sigma: A Systematic Literature Review of the Evidence and Implications. *IEEE Trans. Eng. Manag.* **2024**, *71*, 10261–10294. [CrossRef]
20. Sun, D. Simulation of Mechanical Component Fault Prediction Model Based on Artificial Neural Network Algorithm. In Proceedings of the 2023 International Conference on Internet of Things, Robotics and Distributed Computing, ICIRDC, Rio De Janeiro, Brazil, 29–31 December 2023; pp. 650–654. [CrossRef]
21. Kumari, S.; Tyagi, A.K. Online Social Networking. In *Online Social Networks in Business Frameworks*; Wiley: Hoboken, NJ, USA, 2024; pp. 255–298. ISBN 9781394231126.
22. Jain, R. Towards a Green Revolution: Sustainable Integration of Industry 6.0 Technologies and Smart Banking Services in the UAE. In *Sustainable Innovation for Industry 6.0*; IGI Global: Hershey, PA, USA, 2024; pp. 31–44.
23. Ikram, L.; Afaf, D.; Latifa, E. Beyond Industry 5.0: Tracing the Past, Present, and Future of Digital Industrial Transformations. In Proceedings of the 4th International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET, Fez, Morocco, 16–17 May 2024. [CrossRef]
24. Wan, P.K.; Leirmo, T.L. Human-Centric Zero-Defect Manufacturing: State-of-the-Art Review, Perspectives, and Challenges. *Comput. Ind.* **2023**, *144*, 103792. [CrossRef]

25. Souza, L.; Camboim, K.; Alencar, F. A Systematic Literature Review about Integrating Dependability Attributes, Performability and Sustainability in the Implantation of Cooling Subsystems in Data Center. *J. Supercomput.* **2022**, *78*, 15820–15856. [CrossRef]
26. Seddon, P.B. Where Do Theories Come From? An Inference-to-the-Best-Explanation Theory of Theory Building (IBET). *Commun. Assoc. Inf. Syst.* **2022**, *51*, 32. [CrossRef]
27. Vila-Henninger, L.; Dupuy, C.; Van Ingelgom, V.; Caprioli, M.; Teuber, F.; Pennetreau, D.; Bussi, M.; Le Gall, C. Abductive Coding: Theory Building and Qualitative (Re) Analysis. *Sociol. Methods Res.* **2024**, *53*, 968–1001. [CrossRef]
28. Ulaga, W.; Kleinaltenkamp, M.; Kashyap, V.; Eggert, A. Advancing Marketing Theory and Practice: Guidelines for Crafting Research Propositions. *AMS Rev.* **2021**, *11*, 395–406. [CrossRef]
29. George, J.M.; Jones, G.R. The Role of Time in Theory and Theory Building. *J. Manag.* **2000**, *26*, 657–684. [CrossRef]
30. Sætre, A.S.; Van de Ven, A. Generating Theory by Abduction. *Acad. Manag. Rev.* **2021**, *46*, 684–701. [CrossRef]
31. Kaur, A.; Potdar, V.; Agrawal, H. Industry 4.0 Adoption in Food Supply Chain to Improve Visibility and Operational Efficiency—A Content Analysis. *IEEE Access* **2023**, *11*, 73922–73958. [CrossRef]
32. Núñez-Merino, M.; Maqueira-Marín, J.M.; Moyano-Fuentes, J.; Martínez-Jurado, P.J. Industry 4.0 and Lean Supply Chain Management: Impact on Responsiveness. In *Enabling Technologies for the Successful Deployment of Industry 4.0*; CRC Press: Boca Raton, FL, USA, 2020; pp. 105–128. ISBN 9780429055621.
33. Pandey, A.K.; Daultani, Y.; Pratap, S.; Ip, A.W.H.; Zhou, F. Analyzing Industry 4.0 Adoption Enablers for Supply Chain Flexibility: Impacts on Resilience and Sustainability. *Glob. J. Flex. Syst. Manag.* **2024**, *26*, 1–24. [CrossRef]
34. Nair, S.; Kumar, A.A.; Chacko, E.; Simon, S. Synergizing Humanity and Technology: A Human-Machine Collaboration for Business Sustainability in Industry 5.0. *Stud. Syst. Decis. Control* **2024**, *535*, 105–115. [CrossRef]
35. Maddikunta, P.K.R.; Pham, Q.-V.; Prabadevi, B.; Deepa, N.; Dev, K.; Gadekallu, T.R.; Ruby, R.; Liyanage, M. Industry 5.0: A Survey on Enabling Technologies and Potential Applications. *J. Ind. Inf. Integr.* **2022**, *26*, 100257. [CrossRef]
36. Zizic, M.C.; Mladineo, M.; Gjeldum, N.; Celent, L. From Industry 4.0 towards Industry 5.0: A Review and Analysis of Paradigm Shift for the People, Organization and Technology. *Energies* **2022**, *15*, 5221. [CrossRef]
37. Vijay, A.J.; William, B.N.J.; Haruna, A.A.; Prasad, D.D. Exploring the Synergy of IIoT, AI, and Data Analytics in Industry 6.0. In *Industry 6.0*; CRC Press: Boca Raton, FL, USA, 2024; pp. 1–36. ISBN 9781040128039.
38. Alexa, L.; Pîslaru, M.; Avasilcăi, S. From Industry 4.0 to Industry 5.0—An Overview of European Union Enterprises. In *Sustainability and Innovation in Manufacturing Enterprises: Indicators, Models and Assessment for Industry 5.0*; Draghici, A., Ivascu, L., Eds.; Springer: Singapore, 2022; pp. 221–231. ISBN 978-981-16-7365-8.
39. Almusaed, A.; Yitmen, I.; Almssad, A. Reviewing and Integrating AEC Practices into Industry 6.0: Strategies for Smart and Sustainable Future-Built Environments. *Sustainability* **2023**, *15*, 13464. [CrossRef]
40. Fekrisari, M.; Kantola, J. Integrating Industry 4.0 in Manufacturing: Overcoming Challenges and Optimizing Processes (Case Studies). *TQM J.* **2024**, *36*, 347–370. [CrossRef]
41. Tóth, A.; Nagy, L.; Kennedy, R.; Bohuš, B.; Abonyi, J.; Ruppert, T. The Human-Centric Industry 5.0 Collaboration Architecture. *MethodsX* **2023**, *11*, 102260. [CrossRef]
42. Tyagi, A.K.; Bhatt, P.; Chidambaram, N.; Kumari, S. Artificial Intelligence Empowered Smart Manufacturing for Modern Society. In *Artificial Intelligence—Enabled Digital Twin for Smart Manufacturing*; Wiley: Hoboken, NJ, USA, 2024; pp. 55–83. ISBN 9781394303601.
43. Minz, N.K. Strategic HRM Techniques and Sustainable Innovation for Industry 6.0. In *Sustainable Innovation for Industry 6.0*; Sharma, A., Moses, O., Sharma, R.B., Gupta, S., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 59–73. ISBN 9798369331408.
44. Lykov, A.; Altamirano Cabrera, M.; Konenkov, M.; Serpiva, V.; Fid'ele Gbagbe, K.; Alabbas, A.; Fedoseev, A.; Moreno, L.; Haris Khan, M.; Guo, Z.; et al. Industry 6.0: New Generation of Industry Driven by Generative AI and Swarm of Heterogeneous Robots. *arxiv* **2024**, arXiv:2409.10106. [CrossRef]
45. Das, S.; Pan, T. A Strategic Outline of Industry 6.0: Exploring the Future. *SSRN Electron. J.* **2022**. [CrossRef]
46. Yadav, S.; Rab, S.; Wan, M. Metrology and Sustainability in Industry 6.0: Navigating a New Paradigm. In *Handbook of Quality System, Accreditation and Conformity Assessment*; Bhatnagar, A., Yadav, S., Achanta, V., Harmes-Liedtke, U., Rab, S., Eds.; Springer: Singapore, 2024; pp. 1–31. ISBN 978-981-99-4637-2.
47. Schou, C.; Colli, M.; Berger, U.; Lassen, A.H.; Madsen, O.; Møller, C.; Wæhrens, B.V. Deconstructing Industry 4.0: Defining the Smart Factory. In *Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems*; Andersen, A.-L., Andersen, R., Brunoe, T.D., Larsen, M.S.S., Nielsen, K., Napoleone, A., Kjeldgaard, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 356–363.
48. Beltrami, M.; Orzes, G.; Sarkis, J.; Sartor, M. Industry 4.0 and Sustainability: Towards Conceptualization and Theory. *J. Clean. Prod.* **2021**, *312*, 127733. [CrossRef]
49. Siqin, T.; Choi, T.-M.; Chung, S.-H.; Wen, X. Platform Operations in the Industry 4.0 Era: Recent Advances and the 3As Framework. *IEEE Trans. Eng. Manag.* **2024**, *71*, 1145–1162. [CrossRef]

50. Torn, I.A.R.; Vaneker, T.H.J. Mass Personalization with Industry 4.0 by SMEs: A Concept for Collaborative Networks. *Procedia Manuf.* **2019**, *28*, 135–141. [CrossRef]
51. Baig, M.I.; Yadegaridehkordi, E. Industry 5.0 Applications for Sustainability: A Systematic Review and Future Research Directions. *Sustain. Dev.* **2024**, *32*, 662–681. [CrossRef]
52. Wang, L.; Ranjan, R.; Chen, J.; Benatallah, B. *Cloud Computing: Methodology, Systems, and Applications*, 1st ed.; Wang, L., Ranjan, R., Chen, J., Benatallah, B., Eds.; CRC Press: Boca Raton, FL, USA, 2012; ISBN 9781315217208.
53. Pattanaik, S.; Mohammed, M.; Sood, V. Artificial Intelligence and Machine Learning in Industry 6.0. In *Industry 6.0*; CRC Press: Boca Raton, FL, USA, 2024; pp. 37–54.
54. Zahariev, P.; Hristov, G.; Chaisricharoen, R.; Bencheva, N.; Beloev, I.; Zlatov, N.; Le, C.H.; Georgiev, G.; Kinaneva, D. Industry 4.0 and Beyond—Present Trends, Emerging Solutions and Future Technologies in the Area of the Industrial Automation. In Proceedings of the 2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Chiang-mai, Thailand, 31 January–3 February 2024; pp. 525–528.
55. Gutta, L.M.; Dhamodharan, B.; Dutta, P.K.; Whig, P. AI-Infused Quantum Machine Learning for Enhanced Supply Chain Forecasting. In *Quantum Computing and Supply Chain Management: A New Era of Optimization*; Hassan, A., Bhattacharya, P., Dutta, P.K., Verma, J.P., Kundu, N.K., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 48–63. ISBN 9798369341070.
56. Heilala, J.; Kantola, J. Sustainable Manufacturer Engineering for Industry 6.0. In *Proceedings of the 15th International Conference on Axiomatic Design 2023*; Puik, E., Cochran, D.S., Foley, J.T., Foith-Förster, P., Eds.; Springer: Cham, Switzerland, 2024; pp. 75–87.
57. Doyle-Kent, M.; Kopacek, P. Optimising Human Potential Through Diversity and Inclusion for Industry/Production 4.0, 5.0 and 6.0. In *Proceedings of the Towards Industry 5.0*; Durakbasa, N.M., Gençyılmaz, M.G., Eds.; Springer: Cham, Switzerland, 2023; pp. 267–276.
58. Reddy, M.S.; Reddy, C.K.K.; Hanafiah, M.M. Climate Change Mitigation and Adaptation Strategies Enhanced by Intelligent Systems in Industry 6.0. In *Maintaining a Sustainable World in the Nexus of Environmental Science and AI*; Singh, B., Kaunert, C., Vig, K., Dutta, S., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 201–228. ISBN 9798369363362.
59. Vaseei, M.; Agha, M.N.J. Competitive Advantage in the Supply Chain with an Emphasis on Blockchain Technology. In *Advanced Businesses in Industry 6.0*; Oskounejad, M.M., Nozari, H., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 125–134. ISBN 9798369331088.
60. Maljugić, B.; Ćoćkalo, D.; Bakator, M.; Stanisavljev, S. The Role of the Quality Management Process within Society 5.0. *Societies* **2024**, *14*, 111. [CrossRef]
61. Verma, P.; Kumar, V.; Yalcin, H.; Daim, T. Organizational Architecture of Strategic Entrepreneurial Firms for Digital Transformation: A Bibliometric Analysis. *Technol. Soc.* **2023**, *75*, 102355. [CrossRef]
62. Maimone, F. Organizing for Industry 4.0. In *Do Machines Dream of Electric Workers?* Solari, L., Martinez, M., Braccini, A.M., Lazazzara, A., Eds.; Springer: Cham, Switzerland, 2022; pp. 1–20.
63. Whig, P.; Madavarapu, J.B.; Yathiraju, N.; Thatikonda, R. Managing Knowledge in the Era of Industry 4.0. In *Knowledge Management and Industry Revolution 4.0*; Scrivener Publishing LLC: Beverly, MA USA, 2024; pp. 239–273. ISBN 9781394242641.
64. Tallat, R.; Hawbani, A.; Wang, X.; Al-Dubai, A.; Zhao, L.; Liu, Z.; Min, G.; Zomaya, A.Y.; Alsamhi, S.H. Navigating Industry 5.0: A Survey of Key Enabling Technologies, Trends, Challenges, and Opportunities. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 1080–1126. [CrossRef]
65. Medishetti, S.K.; Karri, G.R.; Pillareddy, V.R.; Sekhar, G.S.; Songa, A.V. IWaOA: Resource Aware Scheduling Through Cloud Fog Computing Environment. In Proceedings of the 2023 4th International Conference on Communication, Computing and Industry 6.0 (C216), Bangalore, India, 15–16 December 2023; IEEE: New York, NY, USA, 2023; pp. 1–6.
66. Tereza, S. Journey to Quality 6.0: Evolution of Quality Management across Industrial Revolutions. *Trends Comput. Sci. Inf. Technol.* **2024**, *9*, 063–070. [CrossRef]
67. Vetrivel, S.C.; Gomathi, T.; Sowmiya, K.C.; Sabareeshwari, V. Customer-Centric Excellence in the Marketing 6.0 Era: Industry 6.0. In *Advanced Businesses in Industry 6.0*; Oskounejad, M.M., Nozari, H., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 192–219. ISBN 9798369331088.
68. Rahmani, R.; Karimi, J.; Resende, P.R.; Abrantes, J.C.C.; Lopes, S.I. Overview of Selective Laser Melting for Industry 5.0: Toward Customizable, Sustainable, and Human-Centric Technologies. *Machines* **2023**, *11*, 522. [CrossRef]
69. Choi, L.; Kim, M.; Kwon, H.-B. Impact of Customer-Based Corporate Reputation on Customer Engagement Behaviors: Customer Identification and Brand Love as Mediators and Industry Type as a Moderator. *J. Brand. Manag.* **2022**, *29*, 150–166. [CrossRef]
70. Villar, A.; Paladini, S.; Buckley, O. Towards Supply Chain 5.0: Redesigning Supply Chains as Resilient, Sustainable, and Human-Centric Systems in a Post-Pandemic World. *Oper. Res. Forum* **2023**, *4*, 60. [CrossRef]
71. Singh, R.; Tyagi, A.K.; Arumugam, S.K. Imagining the Sustainable Future with Industry 6.0: A Smarter Pathway for Modern Society and Manufacturing Industries. In *Machine Learning Algorithms Using Scikit and TensorFlow Environments*; Baby Maruthi, P., Prasad, S., Tyagi, A.K., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 318–331. ISBN 9781668485316.

72. Bosovska, M.; Boiko, M.; Bovsh, L.; Okhrimenko, A. Models of the Industrial Revolution 5.0. In Proceedings of the 2022 IEEE 4th International Conference on Modern Electrical and Energy System (MEES), Kremenchuk, Ukraine, 20–23 October 2022; IEEE: New York, NY, USA, 2023; pp. 1–4.
73. Nozari, H.; Chobar, A.P. The Dimensions and Components of Marketing 5.0: Introduction to Marketing 6.0. In *Advanced Businesses in Industry 6.0*; Oskounejad, M.M., Nozari, H., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 75–86. ISBN 9798369331088.
74. Reddy, A.V.; Doss, S.; Priyanka, K. Human-Machine Collaboration and Emotional Intelligence in Industry 6.0: Concepts, Challenges, and Future Directions. In *Examining the Metaverse in Healthcare: Opportunities, Challenges, and Future Directions*; Nijalingappa, P., Gururaj, T., Goyal, S.B., Shukla, V., Bruno, A., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 221–246. ISBN 9798369315156.
75. Tyagi, A.K.; Kumari, S. Future Opportunities Towards Online Social Networking in the Era of Industry 4.0/5.0. In *Online Social Networks in Business Frameworks*; Wiley: Hoboken, NJ, USA, 2024; pp. 227–254. ISBN 9781394231126.
76. Pattnaik, M.; Pattnaik, A.; Roul, M.K.; Kumar, P.A. Integrated Multi-Item Multi-Echelon Supply Chain Model: A Mixed-Integer Heuristic Approach. In *Evolution and Advances in Computing Technologies for Industry 6.0*; CRC Press: Boca Raton, FL, USA, 2024; pp. 325–376.
77. Dacre, N.; Yan, J.; Frei, R.; Al-Mhdawi, M.K.S.; Dong, H. Advancing Sustainable Manufacturing: A Systematic Exploration of Industry 5.0 Supply Chains for Sustainability, Human-Centricity, and Resilience. *Prod. Plan. Control* **2024**, *36*, 1499–1528. [CrossRef]
78. Annanperä, E.; Jurmu, M.; Kaivo-oja, J.; Kettunen, P.; Knudsen, M.; Lauraéus, T.; Majava, J.; Porras, J. From Industry x to Industry 6.0: Antifragile Manufacturing for People, Planet, and Profit with Passion; 2021. Available online: https://www.businessfinland.fi/4a5d8b/globalassets/julkaisut/industry-x-white-paper.pdf (accessed on 24 August 2025).
79. Mohapatra, S.; Sil, R.; Nag, A. Driving Sustainable Practices in Industry 6.0: A Study on Article Detection with Fuzzy-Wuzzy and Supervised Machine Learning Algorithms. In *Evolution and Advances in Computing Technologies for Industry 6.0*; CRC Press: Boca Raton, FL, USA, 2024; pp. 232–245.
80. Corley, K.G.; Gioia, D.A. Building Theory about Theory Building: What Constitutes a Theoretical Contribution? *Acad. Manag. Rev.* **2011**, *36*, 12–32. [CrossRef]
81. Shepherd, D.A.; Suddaby, R. Theory Building. *J. Manag.* **2017**, *43*, 59–86. [CrossRef]
82. Horváth, I. Theory Building in Experimental Design Research. In *Experimental Design Research*; Cash, P., Stanković, T., Štorga, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 209–231.
83. Luthans, F.; Stewart, T.I. A General Contingency Theory of Management. *Acad. Manag. Rev.* **1977**, *2*, 181–195. [CrossRef]
84. Taher, M. Resource-Based View Theory. In *Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1*; Dwivedi, Y.K., Wade, M.R., Schneberger, S.L., Eds.; Springer: New York, NY, USA, 2012; pp. 151–163.
85. Land, A.; Gruchmann, T.; Siems, E.; Beske-Janssen, P. Dynamic Capabilities Theory. In *Handbook of Theories for Purchasing, Supply Chain and Management Research*; Edward Elgar Publishing: Cheltenham, UK, 2022; pp. 378–398. ISBN 1839104503.
86. Berta, W.; Cranley, L.; Dearing, J.W.; Dogherty, E.J.; Squires, J.E.; Estabrooks, C.A. Why (We Think) Facilitation Works: Insights from Organizational Learning Theory. *Implement. Sci.* **2015**, *10*, 141. [CrossRef]
87. Whitchurch, G.G.; Constantine, L.L. Systems Theory. In *Sourcebook of Family Theories and Methods: A Contextual Approach*; Boss, P., Doherty, W.J., LaRossa, R., Schumm, W.R., Steinmetz, S.K., Eds.; Springer US: Boston, MA, USA, 1993; pp. 325–355. ISBN 978-0-387-85764-0.
88. Yammarino, F.J.; Dubinsky, A.J. Transformational Leadership Theory: Using Levels of Analysis to Determine Boundary Conditions. *Pers. Psychol.* **1994**, *47*, 787–811. [CrossRef]
89. Parmar, B.L.; Freeman, R.E.; Harrison, J.S.; Wicks, A.C.; Purnell, L.; de Colle, S. Stakeholder Theory: The State of the Art. *Acad. Manag. Ann.* **2010**, *4*, 403–445. [CrossRef]
90. Porter, M.E. Technology and Competitive Advantage. *J. Bus. Strategy* **1985**, *5*, 60–78. [CrossRef]
91. Tzortzaki, A.M.; Mihiotis, A. A Review of Knowledge Management Theory and Future Directions. *Knowl. Process Manag.* **2014**, *21*, 29–41. [CrossRef]
92. Di Vaio, A.; Latif, B.; Gunarathne, N.; Gupta, M.; D'Adamo, I. Digitalization and artificial knowledge for accountability in SCM: A systematic literature review. *J. Enterp. Inf. Manag.* **2024**, *37*, 606–672. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

Opportunities for Adapting Data Write Latency in Geo-Distributed Replicas of Multicloud Systems

Olha Kozina [1,*], José Machado [2,*], Maksym Volk [3], Hennadii Heiko [4], Volodymyr Panchenko [4], Mykyta Kozin [3] and Maryna Ivanova [5]

[1] AFOREHAND Studio, 61072 Kharkiv, Ukraine
[2] MEtRICs Research Centre, School of Engineering, University of Minho, Campus of Azurém, 4800-058 Guimarães, Portugal
[3] Department of Electronic Computers, Kharkiv National University of Radio Electronics, 61166 Kharkiv, Ukraine; maksym.volk@nure.ua (M.V.); mykyta.kozin@nure.ua (M.K.)
[4] Computer Engineering and Programming Department, National Technical University Kharkiv Polytechnic Institute, 61002 Kharkiv, Ukraine; hennadii.heiko@khpi.edu.ua (H.H.); volodymyr.panchenko@khpi.edu.ua (V.P.)
[5] Department of Mechanical Engineering Technology and Metal-Cutting Machines, National Technical University Kharkiv Polytechnic Institute, 61002 Kharkiv, Ukraine; maryna.ivanova@khpi.edu.ua
* Correspondence: kozina@aforehand.com.ua (O.K.); jmachado@dem.uminho.pt (J.M.)

Abstract

This paper proposes an AI-based approach to adapting the data write latency in multicloud systems (MCSs) that supports data consistency across geo-distributed replicas of cloud service providers (CSPs). The proposed approach allows for dynamically forming adaptation scenarios based on the proposed model of multi-criteria optimization of data write latency. The generated adaptation scenarios are aimed at maintaining the required data write latency under changes in the intensity of the incoming request flow and network transmission time between replicas in CSPs. To generate adaptation scenarios, the features of the algorithmic Latord method of data consistency, are used. To determine the threshold values and predict the external parameters affecting the data write latency, we propose using learning AI models. An artificial neural network is used to form rules for changing the parameters of the Latord method when the external operating conditions of MCSs change. The features of the Latord method that influence data write latency are demonstrated by the results of simulation experiments on three MCSs with different configurations. To confirm the effectiveness of the developed approach, an adaptation scenario was considered that allows reducing the data write latency by 13% when changing the standard deviation of network transmission time between DCs of MCS.

Keywords: multicloud systems; data writing latency; optimization; data consistency method

1. Introduction

Interest in multicloud systems (MCSs), which make it possible to create and manage copies of data—replicas—in independent geo-distributed databases or storage systems, has been steadily growing [1]. A distinctive feature of MCSs for cloud service customers is the ability to use resources from multiple CSPs, which allows clients to remain independent of the terms of cooperation with any single CSP [2]. In addition, the advantages of MCSs include flexible management of cloud service costs across multiple CSPs, whose data centers (DCs) can be located as close as possible to end users.

When organizing the functioning of MCSs, all issues related to coordinating and managing interactions between services and data hosted in different CSPs fall under the responsibility of cloud service customers [3]. The trade-off between data consistency in MCS replicas and application response latency, incurred during data writes and reads, represents a fundamental principle of any distributed system [4]. At the same time, the development of methods to ensure the relevance and consistency of data within the limits acceptable from the point of view of business logic in each replica of any CSP in a given MCS remains an unresolved scientific and practical problem.

Traditionally, in the literature, the advantages of a new method of data consistency or replication in cloud systems are confirmed by the result of a numerical comparison of write/read delays arising when using the proposed method and a reference method. In most cases, real experiments are conducted on the resources of one CSP, with explanations of the difficulties that may occur when using the obtained results for MCSs. This way of presenting data consistency methods does not allow for estimating the latency that may arise when implementing these methods in other cloud system configurations, especially in MCSs. This significantly reduces the likelihood of the practical use of the presented developments. In addition, it limits the possibility of a priori assessment or control of data write and read latency in the MCS required for switching to another data consistency method.

A reasonable assessment of switching to a new and faster data consistency method in MCSs requires long-term and expensive test implementation and validation, and the process of its implementation may lead to changes in the overall architecture or in the interaction between existing services. Thus, managing data write latency in replicas by switching to new data consistency methods may be difficult to implement in practice. A more realistic option, in this case, is the possibility of initially using such a data consistency method that allows managing data write latency in changing operating conditions for the MCS.

Our main contribution is an approach to adapting the data write latency in geo-distributed MCSs based on the use of AI models and internal parameters of the method Latord of data consistency, in replicas of different CSPs.

Unlike non-algorithmic data consistency methods in cloud systems, the interval parameters of the Latord method determine the data write latency in replicas. The proposed approach to adapting the data write latency in MCSs combines the interval parameters of data reconciliation in replicas, the significance vector of DCs in the MCS, the properties of global communication channels, and incoming request flows into a multi-criteria model.

The results of simulation experiments in three MCSs with different configurations demonstrate a significant effect of the parameters of data reconciliation on individual components of the data write latency, even with a 50-fold increase in the intensity of the incoming write request flow. Using the results of simulation experiments, we propose an example of the formation of adaptation scenarios, the use of which reduces the data write latency in replicas by approximately 13% when changing the properties of global communication channels between DCs of the MCS.

Features of the data consistency the Latord method which define the distinct components of data write latency, the multi-criteria model of the data write optimization, and the approach to the formation of adaptation scenarios on AI models, are presented in Section 3. Section 4 describes the results of simulated experiments. A scenario of the data write latency adaptation based on the experimental results is discussed in Section 5. Section 6 provides our conclusions and outlines the focus of our future work.

2. Related Work

In single-provider cloud systems, the problem of reducing data replication latency can be solved by choosing the optimal number and location of replicas to work with the selected volume of data [5,6]. In geo-distributed systems, when a client requests a read operation, long tail latency may occur, to reduce which a predictive replica selection strategy was developed in [7]. In [8], the authors proved that a machine learning (ML) method can be effective in automatic threshold-based mechanisms of data placement and replication activation.

To mitigate the well-known conflict between the cost and latency of read requests, the geo-distributed cloud storage GeoCol was proposed in [9]. GeoCol integrates AI-based methods for request partitioning and storage planning. GeoCol uses an ML technique to predict request latency and a reinforcement learning (RL) model to ensure that each sub-request is sent in parallel to the optimally located DC of the CSP.

In [10], based on a comparison of various principles for classifying data replication strategies, a conclusion was made about the advisability of using not only the response time to requests but also an assessment of the economic effect as the goal of replication in single-provider cloud systems.

In [11], the authors propose making decisions about the need to create or delete replicas in single-provider cloud systems based on the assessment of the response time to read requests and the costs that affect the profitability of the cloud provider. The paper's approach to calculating the response time to read requests allocates the maintenance of an up-to-date list of replicas in the cloud system to the internal mechanisms of a database, which would be difficult for MCSs.

A dynamic replica placement strategy in MCSs based on predefined service level agreement (SLA) thresholds for the response time to read requests, cost, and data popularity is proposed in [12]. This approach can be very effective in achieving tenant performance objectives, given their economic constraints.

Data processing involves distributed execution and partitioning, even within a single-provider cloud system. SLA-defined limit request response time, as a parameter of the quality of service of one provider, can be used as an inter-region delay only within one CSP. If the query execution plan contains data from replicas in other CSPs, it is necessary to determine the data transfer time between different CSPs, which is not regulated by SLA.

The average backbone network latency between Azure datacenters [13] and the AWS regions' latency matrix [14] demonstrates that there can be a tenfold difference in the average inter-region latency, even for single providers. The average network latency of packet transmission between replicas of different CPSs can differ significantly from the average inter-region latency in one CSP, which must be taken into account when calculating request response time in MCSs.

The strategy of data placement in DCs of various CSPs was analyzed in [15]. The authors demonstrated that placing data as close as possible to the location of data users reduces the response time to read requests by more than 20%, but in this case, it is necessary to replicate and check data consistency across all CSPs.

Various groups of methods were used to ensure data consistency in geo-distributed replicas belonging to different CSPs. The largest group comprised replication protocols that are variations of the Paxos protocol [16–19], differing in the way that a leader replica is chosen.

It is also possible to identify a group of replication protocols in which data consistency is achieved using timestamps and various methods for synchronizing physical and logical clocks [20–22]. However, the data reconciliation methods of this group are effective when

applied to the resources of a single CSP, but they lose their advantages when implemented in the MCSs.

In [23], the VIP-Grab data replication protocol based on an algorithmic method of data consistency in MCSs was proposed. However, principles of the formation of a complete ordered sequence of numbers to data write requests in this protocol can be the cause of long response time to data write requests.

The above methods do not contain parameters whose change could affect the latency of data write and read requests in geo-distributed replicas of different CSPs of the MCS. At the same time, the interval method of ordering data write requests proposed in [24] contains such parameters.

AI models are used to address various problems that arise in cloud services: from adapting infrastructure to traffic times [25] to forecasting workload demand and dynamically managing multiple resources [26,27]. In [28], the authors developed a deep learning algorithm that predicts Round-Trip delay Time (RTT) and the availability of cloud services. Real-time live data from Amazon Web Services, Alibaba Cloud, and Tencent Cloud serve as historical data for the real-time analysis and prediction of RTT between the DCs of CSPs. The article discusses how the research results can help cloud service customers choose CSPs to create more efficient MCSs without reducing the data write latency in them.

As far as we know, AI models have not been applied to adapt data write delay based on the data consistency method in MCSs. Predictions of external factors for MCSs together with the use of our method to coordinate data write requests will allow the data write delay to be adapted to them in advance. We note that our work is novel to the best of our knowledge in that it demonstrates the feasibility of adapting data write latency in MCSs based on a query consistency method and AI models.

3. Materials and Methods

3.1. MCS Configuration

Our goal is to investigate the possibility of managing data write latency in geo-distributed replicas of different CSPs when there are changes in the external operating parameters of the MCS. The interaction between hardware and software resources (servers, virtual machines, databases, and services), as well as the methods of data replication within each CSP, is carried out by means of the CSP itself and can differ significantly, in principle, from each other. Meanwhile, interactions between CSPs are performed via identical instances of the middleware, located in a single DC of each CSP, referred to as a broker Br_x, $x = 1 \ldots BrMax$, $BrMax$—the number of brokers in an MCS. Thus, we consider an MCS as a network of geo-distributed brokers containing data replicas. Each broker serves as the geographical center of a virtual cluster of end users whose requests are directed to it. Figures 1–3 show geographical maps of broker locations as red dots and their corresponding virtual user clusters as red ovals for the three MCSs whose parameters are used in this study.

- MCS A contains brokers in Brazil, South Africa, China, and Malaysia;
- MCS B contains brokers in India, Canada, Israel, and the United Kingdom;
- MCS C contains brokers in China, Japan, Malaysia, and South Korea.

The green lines in these figures depict the conditional connections between brokers, illustrating the distances between them.

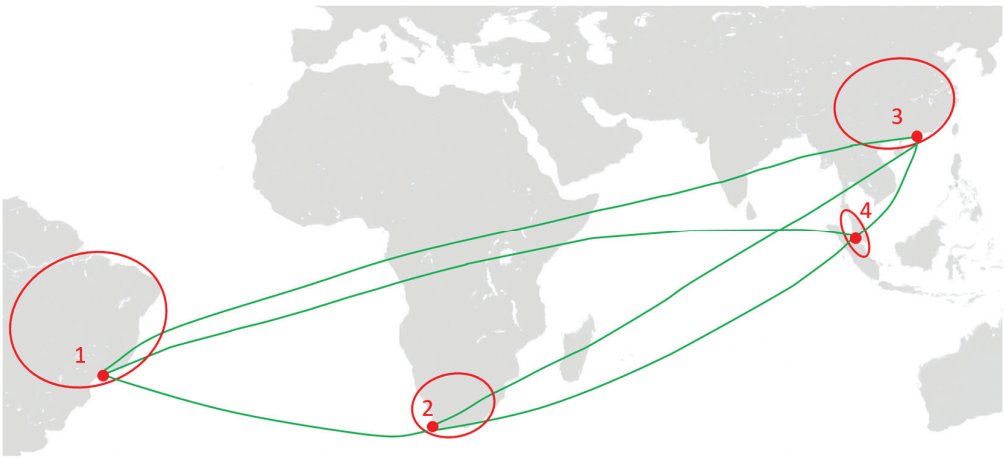


Figure 1. Scheme of the conditional connections (green lines) between brokers (red points) and user clusters (red ovals) for MCS A: 1—broker Br_1 in São Paulo (Brazil), 2—broker Br_2 in Cape Town (South Africa), 3—broker Br_3 in Hong Kong (China), and 4—broker Br_4 in Kuala Lumpur (Malaysia).

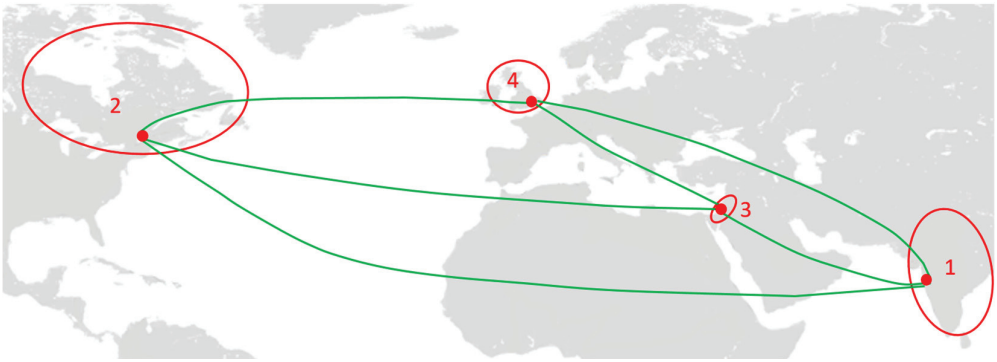


Figure 2. Scheme of the conditional connections (green lines) between brokers (red points) and user clusters (red ovals) for MCS B: 1—broker Br_1 in Hyderabad (India), 2—broker Br_2 in Montreal (Canada), 3—broker Br_3 in Tel Aviv (Israel), and 4—broker Br_4 in London (United Kingdom).

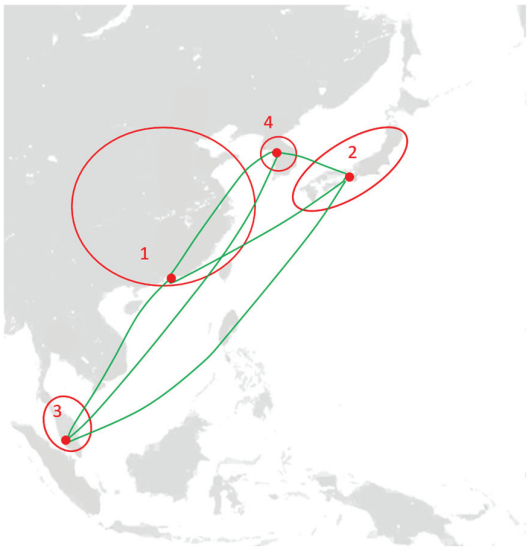


Figure 3. Scheme of the conditional connections (green lines) between brokers (red points) and user clusters (red ovals) for MCS C: 1—broker Br_1 in Hong Kong (China), 2—broker Br_2 in Osaka (Japan), 3—broker Br_3 in Kuala Lumpur (Malaysia), and 4—broker Br_4 in Seoul (South Korea).

3.2. General Principles of Interval-Based Coordination of the Data Write Requests with the Latord Method in Geo-Distributed Replicas

Each broker, upon receiving a data request from a user, appends a key packet $PR(Br_x(to), Ngl(to))$ [29] in the form of a tuple of elements (Table 1) and forwards it to all other brokers via the TCP protocol.

Table 1. Elements of tuple $PR(Br_x(to), Ngl(to))$.

Element of Tuple	Description
$Br_x(from)$	the source broker that received the data packet from the user
$N^W temp(from)$	the sequence number of the current packet among those received by the source broker $Br_x(from)$ within the availability window W_x
$Ntemp(from)$	the sequence number of the current packet among those received by the source broker $Br_x(from)$ within the residual window ΔI_x
$Ngl(from)$	the packet number in the common order sequence at the source broker $Br_x(from)$ at the time of key packet formation
$I(from)$	the number of the adjustment interval during which a packet from a user arrived at the source broker $Br_x(from)$
$N^W temp(to)$	the sequence number of the current packet among those received by the current broker $Br_x(to)$ within the availability window W_x
$Ntemp(to)$	the sequence number of the current packet among those received by broker $Br_x(to)$ within the residual window ΔI_x
$Ngl(to)$	the current packet number in the common order sequence at broker $Br_x(to)$
$I(to)$	the number of the adjustment interval during which the key packet was received by broker $Br_x(to)$

At that moment, a broker $Br_x(from)$ can be considered an active replication node. Upon receiving a key packet from another broker, a broker $Br_x(to)$ is considered a passive node. Thus, each broker functions as both a leader and a follower in MCS replication at different points in time.

It is assumed that all passive nodes receive the packet sent by the active node only once during the specified time interval. In this case, the order in which key packets are received by each broker may not correspond to the order in which these packets were sent by the active node. As a result, each broker eventually receives copies of all requests handled by the MCS, but according to its own observed chronology, i.e., the order of receipt. The main task of the Latord method is to assign, in each broker, specific numbers to all packets (i.e., user requests and key packets), in accordance with the common order sequence within the established acceptable interval. These idempotently consistent numbers determine the order of data writes across all replicas, thereby defining the degree of data consistency in the MCS.

For packet identification by monotonically increasing natural numbers that form the common order sequence, the incoming stream at each broker is treated as a sequence of closed intervals $I(to)$ with a fixed duration. Sorting packets within a closed interval, i.e., sorting a finite subset of packets from an otherwise infinite incoming packet stream, eliminates the problem of sorting elements of an infinite set [30]. Sorting the packet numbers within each interval creates, in each of them, a complete partially ordered set, in which, for any pair of numbers a and b, the condition $a < b$ holds. The Latord method determines the minimum number $Nfree(I_x(to))$ of the first packet received by the broker in the current adjustment interval $I_x(to)$ as one greater than the maximum packet number in the sequence formed in the previous interval $I_x(to - 1)$. Such a linear alignment of numbers between

adjacent intervals makes it possible to treat the tuple of ordered packet numbers from the incoming stream as a total order sequence.

The correct number of the current packet in a common order sequence is determined immediately upon its arrival at any broker. This means that, from all packets previously received by a given broker, a common order sequence has already been formed, in which the correct position for the current packet must be determined. The rules for determining the correct position, i.e., the number in a common order sequence, of the current packet are implemented in our Corr algorithm. In this case, inserting the current packet into the correct position within the sequence of numbers results in the adjustment of the numbers previously assigned to earlier packets, whereas adding the current packet with a new maximum number in the sequence does not.

Accordingly, the use of the Latord method for interval-based coordination of data write requests at each broker, for any packet received within the current interval, can be divided into two consecutive steps:

1. Formation of the tuple *Mlate* of prior time intervals, whose packets may be adjusted, and the tuple *Mcor* of packet numbers previously received by this broker from *Mlate*;
2. Sorting the packet numbers from *Mcor*, with the consideration of the parameters of the currently received packet, using the Corr algorithm.

3.2.1. Formation of the *Mlate* Time Interval Tuple and the *Mcor* Packet Tuple

In the MCS, global network channels are used for packet transmission, both between users and brokers and among brokers themselves, with uncontrolled fluctuations in signal transmission speed. This implies that all time intervals describing packet transmission should be treated as random variables.

For each broker Br_x, the interval I_x, in which packet numbers are sorted, is represented as a tuple consisting of the availability window W_x and the residual window ΔI_x. The duration of the sorting interval equals the sum of the corresponding windows [24]:

$$I_x = W_x + \Delta I_x. \tag{1}$$

The duration of the availability window W_x depends on the area of the notional user cluster of the current broker Br_x and can be estimated numerically as the average time required for a request to reach the corresponding nearest broker [24]. This implies that for a notional cluster of active MCS users distributed over a larger geographical area, the availability window must be larger than for a cluster with a more compact distribution of active users. The duration of the residual window ΔI_x should be no less than the average time of packet transmission between the current broker and the most distant broker within the given MCS [24].

The sorting interval I_x differs across brokers; therefore, a single value, referred to as the adjustment interval I, is selected for all brokers to synchronize the packet coordination process in the MCS [24]:

$$I = \max_{x=1\ldots BrMax} I_x. \tag{2}$$

Thus, the time axis is represented as a sequence of non-overlapping adjustment intervals I, consecutively numbered with natural numbers. Meanwhile, the adjustment interval I at each broker preserves its structure, remaining a tuple consisting of the availability window W_x, defined for each broker, and the residual window ΔI_x.

Using a global network to transmit packets between brokers may result in a key packet being sent from interval $I(from)$ and arriving at another broker after the end of the same interval $I(from)$. In other words, a packet may "arrive late" relative to the end of the interval from which it was sent and will be received during the following adjustment intervals $I(to)$.

We will consider it a delay when a single key packet contains $I(to) > I(from)$. The maximum allowable packet delay, *MaxLate*, is determined by the following relationship [24]:

$$MaxLate = K_{Late} \cdot I, \tag{3}$$

where K_{Late} is a positive integer. For example, if a packet is sent by a broker from window W_x of interval $I(to)$, and *MaxLate* equals $2 \cdot I$, then the latest, i.e., the final, window in which this packet can be received by another broker will be W_x of interval $I(to + 2)$.

The value of *MaxLate* is used in the Latord method to determine the order of packet transmission without relying on timestamps. Knowing the window type and the interval number of the current packet's arrival, for example, ΔI_x from $I(to)$, and assuming that the packet is maximally delayed, we consider the source of this packet to be the corresponding window ΔI_x of interval $I(to - K_{Late})$. If this assumption is correct, the number of the current packet should be smaller than the number of all previously received packets that were sent by other brokers after the current packet was sent. In other words, the number of the current packet should be smaller than the number of previously received packets that were sent from the tuple of windows *Mlate*, ranging from ΔI_x of interval $I(to - K_{Late})$ to ΔI_x of interval $I(to)$. Such a tuple of windows, *Mlate*, will contain $2 \cdot K_{Late} + 1$ windows, while the tuple *Mcor* will include all packets that have already arrived at the current broker from these $2 \cdot K_{Late} + 1$ windows. Representing the arrays *Mlate* and *Mcor* as tuples emphasizes the need to arrange both the windows and the packets within them in ascending order of their previously assigned sequence numbers. The hypothesis regarding the transmission order of the current packet among the packets in the tuple *Mcor* is verified according to the rules of the Corr sorting algorithm.

Let us consider an example of the formation of the tuples *Mlate* and *Mcor* in the MCS consisting of three brokers, for which the maximum allowable packet delay *MaxLate* is $2 \cdot I$. Let the current packet, marked in red in Figure 4, have arrived at broker Br_2 during availability window ΔI_2 of adjustment interval $I(57)$.

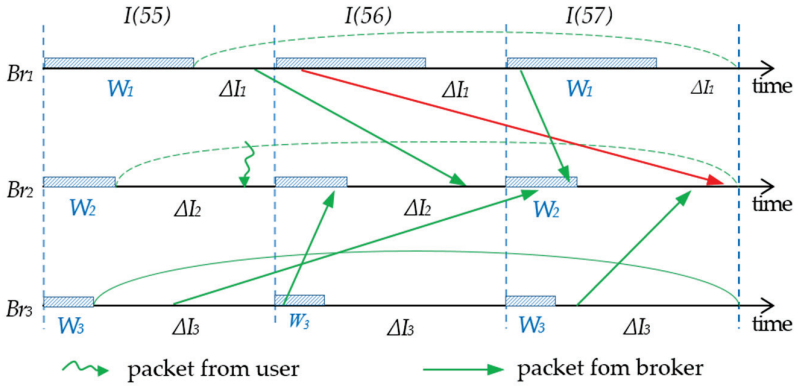


Figure 4. Example of the formation of the *Mlate* window tuple and the *Mcor* packet tuple for the current packet (red arrow).

For such a packet, the *Mlate* window tuple contains 5 elements:

$$Mlate = \langle \Delta I_2(55), W_2(56), \Delta I_2(56), W_2(57), \Delta I_2(57) \rangle, \tag{4}$$

while the *Mcor* packet tuple contains 6 packets previously received by Br_2, marked in green in Figure 4. The *Mcor* tuple contains both packets sent by other brokers from the windows of the *Mlate* tuple and packets received from users during the same windows.

3.2.2. The Corr Algorithm for Determining the Correct Packet Number in the Common Order Sequence of Numbers

The sorting conditions according to the Corr algorithm are checked between the currently received packet and each packet in the *Mcor* tuple by iterating through all elements of *Mcor* sequentially, starting from the first.

In the Corr algorithm, the conditions for the pairwise agreement of packet numbers correspond not to the actual chronology of packet arrivals at the brokers, but rather to the chronology of the windows in the *Mlate* tuple. This implies that all packets sent from earlier windows must be assigned smaller numbers in the sequence than those sent from later windows. For requests received from users in a key packet, the broker that received the request will be considered the source broker of the packet; that is, $Br_x(from)$ and $Br_x(to)$ in the key packet will be equal.

When comparing packets sent from the same window of a given adjustment interval but originating from different brokers, the packet from the broker with a higher priority is assigned the smaller sequence number. A broker priority corresponds to the duration of its availability window W_x: a broker with a longer W_x is assigned a higher service priority than a broker with a shorter W_x [24]. This allows packets sent by users who are theoretically farther from their broker in larger clusters to be "moved ahead" of packets sent by users who are closer to the broker in smaller clusters.

When comparing packets sent by the same broker from a single window W_x or ΔI_x, the packet sent earlier—that is, the packet with the smaller number $N^W temp(from)$ or $Ntemp(from)$, respectively—receives the smaller sequence number.

3.3. Data Write Latency in the MCS Implementing the Latord Method

The maximum allowable interval, *MaxLate*, for packet delivery between brokers effectively defines the period during which the number of the current packet can be adjusted according to the parameters of subsequent packets. After $2 \cdot K_{Late}$ subsequent windows have passed, the current packet will no longer be included in the *Mcor* tuples created for sorting new packets. This indicates that after $2 \cdot K_{Late}$ subsequent windows have elapsed, the packet number $Ngl(to)$ in the total order sequence can be considered finally adjusted and used to establish the order of writing the corresponding user data to all replicas. We will consider $LatencyWr(Ngl(to))$ as the delay in writing data to the replica—the time interval between the moment it becomes impossible to adjust the $Ngl(to)$ number of the packet and the moment the corresponding request from the user reaches the active broker $Br_x(from)$:

$$LatencyWr(Ngl(to)) = \Delta r(Ngl(to)) + MaxLate = \Delta r(Ngl(to)) + K_{Late} \cdot I, \quad (5)$$

where $\Delta r(Ngl(to))$ is the time interval from the moment a user request arrives until the end of the precise incoming window—W_x or ΔI_x. Thus, the latency $LatencyWr(Ngl(to))$ of writing data when using the Latord method is determined by the maximum allowable packet delay *MaxLate*, which depends on the adjustment interval *I*, as well as the proportion of the window W_x to the window ΔI_x in each broker. The component $\Delta r(Ngl(to))$ contributes to the random nature of data write latency. Moreover, the procedure for writing data from a packet with the number $Ngl(to)$ into each replica must begin immediately after the end of the interval defined by expression (5), provided that the packet with the number $Ngl(to)-1$ has already been written onto the current replica. In each replica, the difference between the moment of receiving a packet with the number $Ngl(to)$ and the start of the procedure for writing data from it, even if the packet $Ngl(to) - 1$ has already been written earlier, is the SLA-limited available time of the corresponding resource. Available time, as a parameter of the quality of service of the database's request response time, is included in

the SLA of each CSP, which makes up the MCS and requires the presence of mechanisms for automatic verification of this productivity metric [31,32]. Confirmation of the quality of service guaranteed in the SLA is a well-known non-trivial task for MCSs [33,34] and does not depend on the method of data reconciliation in the replicas.

3.3.1. Parameters of the Latord Method for Assigning Data Write Latency in the MCS

We will consider the latency $Latency_x(Ngl(to))$ of forming the correct number $Ngl(to)$ of the packet at broker Br_x as the interval between the moment $t_x^{cor}(Ngl(to))$ of the last adjustment of the packet number $Ngl(to)$ and the moment $t_x^{in}(Ngl(to))$ this packet arrives at Br_x:

$$Latency_x(Ngl(to)) = t_x^{cor}(Ngl(to)) - t_x^{in}(Ngl(to)). \quad (6)$$

The number of sorts for each packet in any broker equals the number of packets that arrive at that broker after the current packet within the current window and in the subsequent $2 \cdot K_{Late}$ windows. Considering that not every sorting operation may lead to an adjustment of a previously assigned packet number, the moment $t_x^{cor}(Ngl(to))$ of the last adjustment of the number $Ngl(to)$ is a random variable. The moment t_x^{in} when a key packet arrives at the current broker Br_x, as well as the moment t_x^{in} when a request arrives from a user, is a random variable and depends on the geographical location of the brokers. Consequently, in expression (6), both terms are determined by random events, and $Latency_x(Ngl(to))$ should also be considered a random variable, whose parameters are defined by the arrangement of brokers in the MCS.

Let us consider the principle of forming the latency $LatencyWr(Ngl(to))$ of writing data and the latency $Latency_x(Ngl(to))$ of forming the correct number $Ngl(to)$ of the packet on the MCS, as illustrated in the example in Figure 4. Suppose a user request arrives at broker Br_1 at moment t_1^{in} during the availability window W_1 of adjustment interval $I(14)$ (Figure 5).

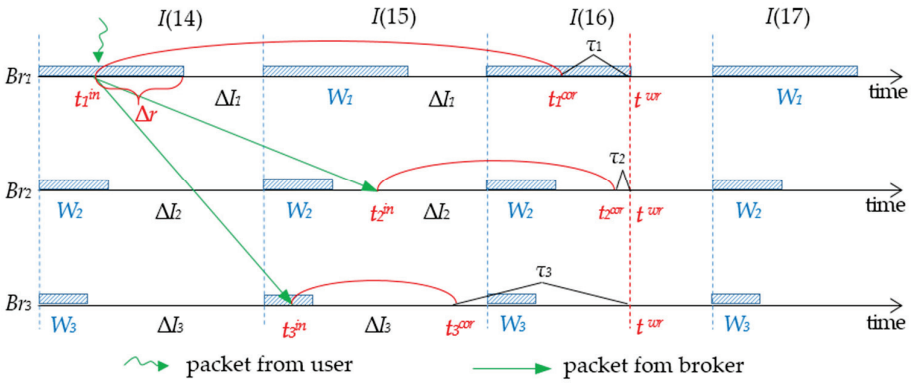


Figure 5. Example of the determination of the latency of forming the correct packet number and the data write latency.

According to the chosen *MaxLate* from expression (5), writing a packet to all replicas can begin at $t^{wr}(Ngl(to))$, i.e., through $2 \cdot I$ after the completion of W_1 from $I(14)$. Therefore, the data write latency for the current packet to all replicas will be equal to

$$LatencyWr(Ngl(to)) = t^{wr}(Ngl(to)) - t_1^{in} = \Delta r(Ngl(to)) + MaxLate = \Delta r(Ngl(to)) + 2 \cdot I. \quad (7)$$

Let us assume that the key packet corresponding to the user's input request arrived at the other brokers at moments t_2^{in} and t_3^{in}. Suppose that the number of sorting procedures for the packet number $Ngl(to)$ in the brokers was different, so that the adjustments were performed independently and completed at different moments $t_x^{cor}(Ngl(to))$, $x = 1 \ldots 3$. Then,

the numerical values of the latencies $Latency_x(Ngl(to))$ in generating the correct number $Ngl(to)$ of the packet in the brokers differ and equal, accordingly, as follows:

$$\begin{array}{l} Latency_1(Ngl(to)) = t_1^{cor}(Ngl(to)) - t_1^{in}, \\ Latency_2(Ngl(to)) = t_2^{cor}(Ngl(to)) - t_2^{in}, \\ Latency_3(Ngl(to)) = t_3^{cor}(Ngl(to)) - t_3^{in}. \end{array} \quad (8)$$

Generally, the data write latency $LatencyWr(Ngl(to))$ can be expressed using the varying values of the latency $Latency_x(Ngl(to))$ for generating the packet number $Ngl(to)$ at different brokers:

$$LatencyWr(Ngl(to)) = \Delta t_{from-to}^{tr}(Ngl(to)) + Latency_x(Ngl(to)) + \tau_x, x = 1 \ldots BrMax, \quad (9)$$

where $\Delta t_{from-to}^{tr}(Ngl(to)) = t_{to}^{in}(Ngl(to)) - t_{from}^{in}(Ngl(to))$ is the packet $Ngl(to)$ transmission duration from the active broker $Br(from)$ to the current broker $Br(to)$, and τ_x is the interval between the data write moment $t^{wr}(Ngl(to))$ and the moment $t_x^{cor}(Ngl(to))$ of the last adjustments of the packet $Ngl(to)$. It is assumed that for a user request at broker Br_1 (Figure 5),

$$\Delta t_{from-to}^{tr}(Ngl(to)) = \Delta t_{1-1}^{tr}(Ngl(to)) = 0. \quad (10)$$

The interval τ_x is an excessive ballast random component of the data write latency $LatencyWr(Ngl(to))$, since not all brokers perform the adjustment of the packet number $N_{gl}(to)$ throughout the entire interval from t_x^{in} to $t^{wr}(N_{gl}(to))$. Therefore, changes in the data recording delay, expressed by Formula (9), can be achieved by changing the duration of packet transmission between brokers $\Delta t_{from-to}^{tr}(Ngl(to))$ and $Latency_x(Ngl(to))$ of the formation of the correct packet number $Ngl(to)$ in each broker.

3.3.2. Parameters That Determine the Latency of Forming the Correct Packet Number in Brokers

The number of sorts for the packet number $Ngl(to)$ at each broker is determined by the number of incoming packets during the fixed interval $MaxLate$ and, therefore, depends on the intensity of the incoming user request stream $UtoBr_x$ at each broker. In turn, the number of sorts for the packet number in the interval from t_x^{in} to $t^{wr}(Ngl(to))$ affects the latency $Latency_x(Ngl(to))$ when generating the $Ngl(to)$ number of this packet. The closer these moments are to each other, the smaller $Latency_x(Ngl(to))$ becomes. The position of $t^{wr}(Ngl(to))$ on the time axis depends on the a priori chosen $MaxLate$ value; so, the latency $Latency_x(Ngl(to))$ is determined by the duration of the adjustment interval I.

The initial selection of the adjustment interval I, as well as subsequent modifications of its value during the operation of the MCS, leads to the problem of revising the relationship between the W_x and ΔI_x windows in each broker, i.e., the problem of secondary selection of these windows within I for all brokers.

It is also possible that an inverse case, when revising the window durations, will lead to a change in the adjustment interval I, thereby affecting both the latency $Latency_x(Ngl(to))$ and the data write latency.

It should be noted that difficulties in synchronizing clocks between the devices of all active MCS users and the brokers of the corresponding clusters may turn the selection of W_x based on real measurements into an empirical problem requiring ad-hoc solutions. Meanwhile, the packet transmission time between brokers, as a random variable, has a minimum limit determined by the geographical locations of the corresponding DCs of CSPs and can serve as a starting point for the secondary selection of windows. Moreover, packet transport between brokers is strongly affected by fluctuations in Internet traffic and global communication channel parameters.

Thus, the latency of generating packet numbers in each broker and write data latency are influenced by the set of internal parameters of the Latord method:

$$U(t) = \{W_x(t), \Delta I_x(t), I_x(t), MaxLate\}, x = 1 \dots BrMax, \quad (11)$$

and by external parameters—the intensity of the incoming user request flow *UtoBr* and the duration of packet transmission $\Delta t^{tr}_{from-to}$ between brokers of the MCS.

3.4. Objective Function for Optimizing Data Write Latency in the MCS

This study assumes that information about the forecasts of fluctuations in external operating parameters of the MCS and their influence on the latency of generating packet numbers will make it possible to create a mechanism for adapting the data write latency to such fluctuations.

We define the optimal data write latency in an MCS as the set $Lat(t)$, in which the minimum latency may be achieved not across all brokers, but only within the most significant ones. The numerical significance values $Signif_x$ of a broker can be determined based on the number of active users or the business value of the corresponding user cluster and can be represented as a vector $[Signif_x]$. Hence, the data write latency is the set of elements:

$$Lat(t) = \left\{ LatencyWr_x\left(UtoBr_x(t), \Delta t^{tr}_{from-to}(t), t\right) \mid [Signif_x] \right\}, x = 1 \dots BrMax\,. \quad (12)$$

Expression (11) defines a set of internal parameters whose changes at time t can lead to an optimal value of data write delay in each broker.

Expressions (1) and (2) serve as constraints on the adjustment interval, similar to the probability $P(K_x)$ of receiving packets by a passive broker after the end of the *MaxLate* period:

$$\mathrm{P}(K_x) \le Up_x\,, x = 1 \dots BrMax, \quad (13)$$

where Up_x denotes the maximum allowable probability of the occurrence of "late" packets in each broker Br_x. The initial selection of the residual window $\Delta I_x(t_0)$ for each broker must satisfy the following condition:

$$\Delta I_x(t_0) \le \max mean\left(\Delta t^{tr}_{x-to}\right), x = 1 \dots BrMax\,. \quad (14)$$

Accordingly, the objective function for the multi-objective optimization of data write latency in the MCS can be defined as:

$$\min Lat(t), \begin{cases} U(t) = \{W_x(t), \Delta I_x(t), I_x(t), MaxLate\} \\ \Delta I_x(t_0) \le \max mean\left(\Delta t^{tr}_{x-to}\right) \\ I(t_0) = \max\limits_{x=1\dots BrMax} I_x(t_0) \\ I_x(t) = W_x(t) + \Delta I_x(t) \\ \mathrm{P}(K_x) \le Up_x \\ x = 1 \dots BrMax \end{cases}. \quad (15)$$

Any MCS can be considered a dynamic system subject to unstable external influences, including flows of incoming user requests and random disturbances on global communication channels. Fluctuations in the external parameters of *UtoBr* and $\Delta t^{tr}_{from-to}$ may be caused by reasons independent of the MCS operation. The level of these fluctuations may be related to the day of the week, season, fashion, and special events in one or more user clusters.

3.5. Formation of Adaptation Scenarios of Data Write Latency in the MCS

To respond promptly to incoming request flow fluctuations, reliable forecasts of such changes are required. For this purpose, AI technologies employing machine learning for real-time data processing can be utilized. AI models should provide dynamic predictions of changes in the incoming request flow for each broker separately, based on which decisions regarding the feasibility and magnitude of adjustments in $U(t)$ will be made. At the same time, to make a decision on how to change the elements of the set $U(t)$ to correct changes in $UtoBr$, additional relevant data are required on:

- What magnitude of change in the incoming request flow rate should be considered significant for triggering adjustments;
- Which changes and in which broker, given the current configuration of variations in incoming request flow intensity, should be used to initiate adjustments? For example, consider a situation where, in the broker with the highest significance, a significant increase in input flow intensity is predicted, but it is twice as small as a predicted significant increase for the same time in a broker with lower significance. In such a case, it is not necessary to make an additional decision regarding which of the two significant predicted values of $U(t)$ should be applied;
- What duration of predicted significant changes in the intensity of the incoming request flow should be considered sufficient to trigger adjustments;
- What magnitude of changes in user activity should prompt a revision of the significance vector $[Signif_x]$?

Such data constitute a set of initial constraints or minimum gradients $Y(t)$, which, when exceeded, trigger decision-making procedures regarding changes in $U(t)$. The elements of $Y(t)$ may also vary depending on current user activity trends, requiring decisions regarding their determination based on AI models employing machine learning (ML) techniques.

A similar situation occurs with the second external component, the random nature of which affects the data write latency—$\Delta t^{tr}_{from-to}$. In the conducted simulation experiments, it was assumed that the packet transmission time between brokers can be characterized by a single probability distribution law and a single standard deviation (SD) value. In this case, the SD of $\Delta t^{tr}_{from-to}$ was adjusted to the same level simultaneously across all brokers. Under real conditions, changes in the mean value and SD of $\Delta t^{tr}_{from-to}$ will occur between one or several pairs of brokers.

Changes in the parameters of the random variable $\Delta t^{tr}_{from-to}$ can be predicted if their causes are associated with variations in user activity within clusters with limited communication line capacities, rather than with emergency events. To make decisions regarding adjustments to the values of $U(t)$, which compensate for changes in packet transmission durations between brokers, additional up-to-date data are required on:

- What magnitude of change in the average and SD of $\Delta t^{tr}_{from-to}$ should be considered significant for triggering adjustments;
- What duration of predicted significant changes in the average and SD of $\Delta t^{tr}_{from-to}$ should be considered sufficient to trigger adjustments.

These data also form part of the set of gradients $Y(t)$.

Thus, to determine the magnitude and specific elements of $U(t)$ from Equation (11), which should be adjusted, it is necessary to employ an AI model for the real-time prediction of changes in the incoming request traffic and parameters of global communication channels. The output of this AI model will serve as the input for another tool, the AI classification model, which is responsible for categorizing current gradients $Y(t)$. In this model, simulation methods or neural networks trained on historical data from a specific MCS can be employed to determine the minimum gradients within the set $Y(t)$.

To solve the optimization problem described by expression (15), it is not enough to use simple algorithms based on "if–then" instructions [35]. The set of decisions regarding changes in $U(t)$ will be called a scenario for adapting the data write latency. Adaptation scenarios are customized for each MCS, depending on the specific conditions of its operation. To form effective scenarios for adapting the data recording delay, we can use a neural network model, the input data for which are the output data of the previously described models. The interaction diagram among these three AI models is shown in Figure 6.

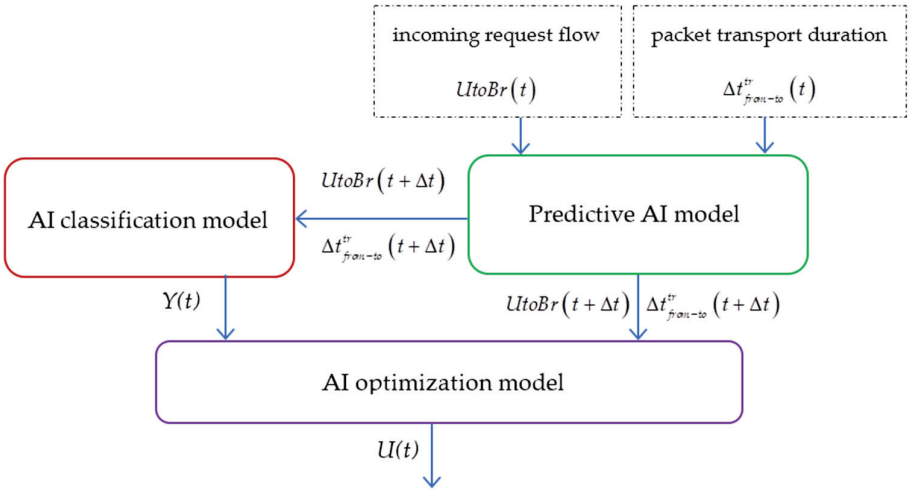


Figure 6. Interaction between AI models for minimizing the data write latency in the MCS.

4. Experimental Setup and Results

The application of the reasonably selected parameters of the Latord method of data consistency, namely, I, W_x, and ΔI_x, will allow for adapting the data write latency $LatencyWr(Ngl(to))$ to changes in the external parameters $UtoBr$ and $\Delta t^{tr}_{from-to}$. In accordance with expressions (15), (12), and (9), the formation of scenarios for adapting the data write latency in MCSs is based on the influence of the elements of the set $U(t)$ on the latency $Latency_x(Ngl(to))$ at different values of the external parameters $UtoBr$ and $\Delta t^{tr}_{from-to}$. To confirm this hypothesis, two groups of simulation experiments were carried out.

The sections above show that the latency $Latency_x(Ngl(to))$ of forming the correct packet number as a separate term in determining the data write latency $LatencyWr(Ngl(to))$ according to Formula (9) depends both on the adjustment interval I and on the ratio of the windows W_x and ΔI_x within it, as given in expression (1). The study of the effect of decreasing I on the latency $Latency_x(Ngl(to))$ of the packet number formation can be carried out in two ways:

- By decreasing the residual window ΔI_x with a fixed availability window W_x. This type of change in the adjustment interval I was used in the experiments of the "varΔI" group;
- By decreasing the availability window W_x with a fixed residual window ΔI_x. This type of change in the adjustment interval I was used in the experiments of the "varW" group.

4.1. MCS Configurations for Experiments

To study the influence of the duration of packet transmission between brokers on the delay $Latency_x(Ngl(to))$, independently of other terms that determine the data write latency $LatencyWr(Ngl(to))$ according to expression (9), simulation experiments were conducted in three MCSs: A, B, and C.

Each MCS consisted of four brokers ($BrMax$ = 4), whose geographical locations are shown in Figures 1–3 and remain unchanged throughout this study. Based on the defined user clusters, with boundaries marked by red ovals in Figures 1–3, the availability window values were selected for each MCS (Table 2).

Table 2. Availability windows for the MCS brokers A, B, and C.

MCS	Availability Window W_x (ms)			
	W_1	W_2	W_3	W_4
A	169.87	69.77	21.75	10.94
B	35.09	27.34	11.09	10.73
C	20.00	14.00	12.00	10.00

Studies [36–38] show that the characteristics of the duration of packet transmission on the Internet depend not only on the distance between the final destinations of the packets but also on the region, country, continent, and Internet user density. The duration of packet transmission over global communication channels in any region of the globe can be described as a random variable that takes on all possible discrete values in a certain range with equal probability. So, in simulation experiments, the packet transmission time $\Delta t^{tr}_{from-to}$ between all pairs of brokers in all MCSs was considered a random variable with a uniform probability distribution. The selection of numerical values of the average $\Delta t^{tr}_{from-to}$ between brokers in MCSs A, B, and C (Table 3) was carried out by taking into account the average latency between Azure DCs on the backbone network [13] and the AWS region latency matrix [14].

Table 3. Mean of packet transmission time between brokers in MCS A, B, and C.

MCS	Brokers	Average of Packet Transmission Between Brokers (ms)			
		Br_1	Br_2	Br_3	Br_4
A	Br_1	4.83	339.73	319.73	334.88
	Br_2	339.45	4.39	242.35	279.81
	Br_3	318.30	241.30	3.43	43.73
	Br_4	337.13	279.08	42.69	3.96
B	Br_1	8.50	221.36	167.63	138.79
	Br_2	226.44	5.11	159.44	92.95
	Br_3	175.47	155.75	4.61	71.49
	Br_4	139.76	91.11	73.95	4.30
C	Br_1	3.43	49.51	43.73	39.94
	Br_2	51.01	4.82	84.57	29.60
	Br_3	44.75	85.85	3.96	80.65
	Br_4	41.78	27.51	79.98	5.59

The average values of $\Delta t^{tr}_{from-to}$ used in all simulation experiments were kept constant.

The long-tailed Pareto distribution, as a special case of the power-law distribution, is widely used in economics, finance, computer science, and the social sciences to describe time intervals between random events [39–41]. So, in simulation experiments, the inter-arrival times of user requests were modeled as independent random variable streams, following a Pareto distribution, with equal intensity across all clusters within each MCS. In this case, the average values of the incoming request flow intensity $UtoBr$ arriving at each broker were 20.00, 50.00, 100.00, and 500.00 requests per second.

For each MCS, simulation experiments of the "varΔI" and "varW" groups were conducted, as described below.

All experiments were implemented using MATLAB R2023b, a platform for computation, data analysis, and visualization.

4.2. Group of Experiments "varΔI"

In the experiments of the "varΔI" group, the effect of changing the adjustment interval I on the latency of forming the correct packet number in each broker of each MCS was investigated. In addition, different intensities of *UtoBr* were used to study the adjustment interval I as the independent term determining the data write latency according to expression (9). In this case, the changes in I were performed by changing the term ΔI_x with a fixed W_x. This situation may arise when the parameters of packet transmission between brokers and from users to brokers, i.e., $\Delta t^{tr}_{from-to}$ and W_x, are defined and considered stable, and the task is to select the optimal I, at which the data write latency will have the required value.

4.2.1. Setup of Experiments "varΔI" in the MCSs

A distinctive feature of the experiments of the "varΔI" group is that the values of the availability windows W_x for brokers of each MCS, given in Table 2, did not change. At the same time, the values of the adjustment interval I_x, $x = 1 \ldots 17$, changed into a wide range of values, the boundary values of which are given in Table 4.

Table 4. Range of changes of the adjustment interval I_x (ms) for the experiments "varΔI" in MCS A, B, and C.

MCS	Range of Changes of I_x (ms)
A	[169.900 . . . 489.900]
B	[113.250 . . . 265.250]
C	[53.125 . . . 108.525]

The changes in the duration of the packet transmission between brokers in the simulation experiments were achieved by using different values of its SD with constant average values (Table 3). For each MCS, two SD variants were used: variant SD1 with basic values and variant SD2 with values that were at least 10 times higher than the values of variant SD1 (Table 5).

Table 5. SD values of the duration $\Delta t^{tr}_{from-to}$ of packet transmission between brokers in the MCSs.

MCS	Brokers	Standard Deviation SD1				Standard Deviation SD2			
		Br_1	Br_2	Br_3	Br_4	Br_1	Br_2	Br_3	Br_4
A	Br_1	0	3.9229	3.6919	3.8669	0	49.0358	46.1491	48.3358
	Br_2	3.9196	0	2.7984	3.2310	48.9954	0	34.9802	40.3871
	Br_3	3.6754	2.7863	0	0.5050	45.9426	34.8287	0	6.3119
	Br_4	3.8928	3.2225	0.4929	0	48.6605	40.2817	6.1618	0
B	Br_1	0	2.5560	1.9356	1.6026	0	31.9506	24.1953	20.0326
	Br_2	2.6147	0	1.8411	1.0733	32.6838	0	23.0132	13.4162
	Br_3	2.0262	1.7984	0	0.8255	25.3269	22.4806	0	10.3187
	Br_4	1.6138	1.0520	0.8539	0	20.1726	13.1506	10.6738	0
C	Br_1	0	0.5717	0.5050	0.4612	0	7.1462	6.3119	5.7648
	Br_2	0.5890	0	0.9765	0.3418	7.3627	0	12.2066	4.2724
	Br_3	0.5167	0.9913	0	0.9313	6.4591	12.3914	0	11.6408
	Br_4	0.4824	0.3177	0.9235	0	6.0304	3.9707	11.5441	0

In each experiment, the incoming flow to each broker consisted of 10,000 user requests for data write.

4.2.2. Results of Experiments "varΔI"

The mean of the packet number formation latencies with SD2 in each broker of MCS A, B, and C form the upper surface in Figures 7–9, whereas with SD1, they form the lower surface.

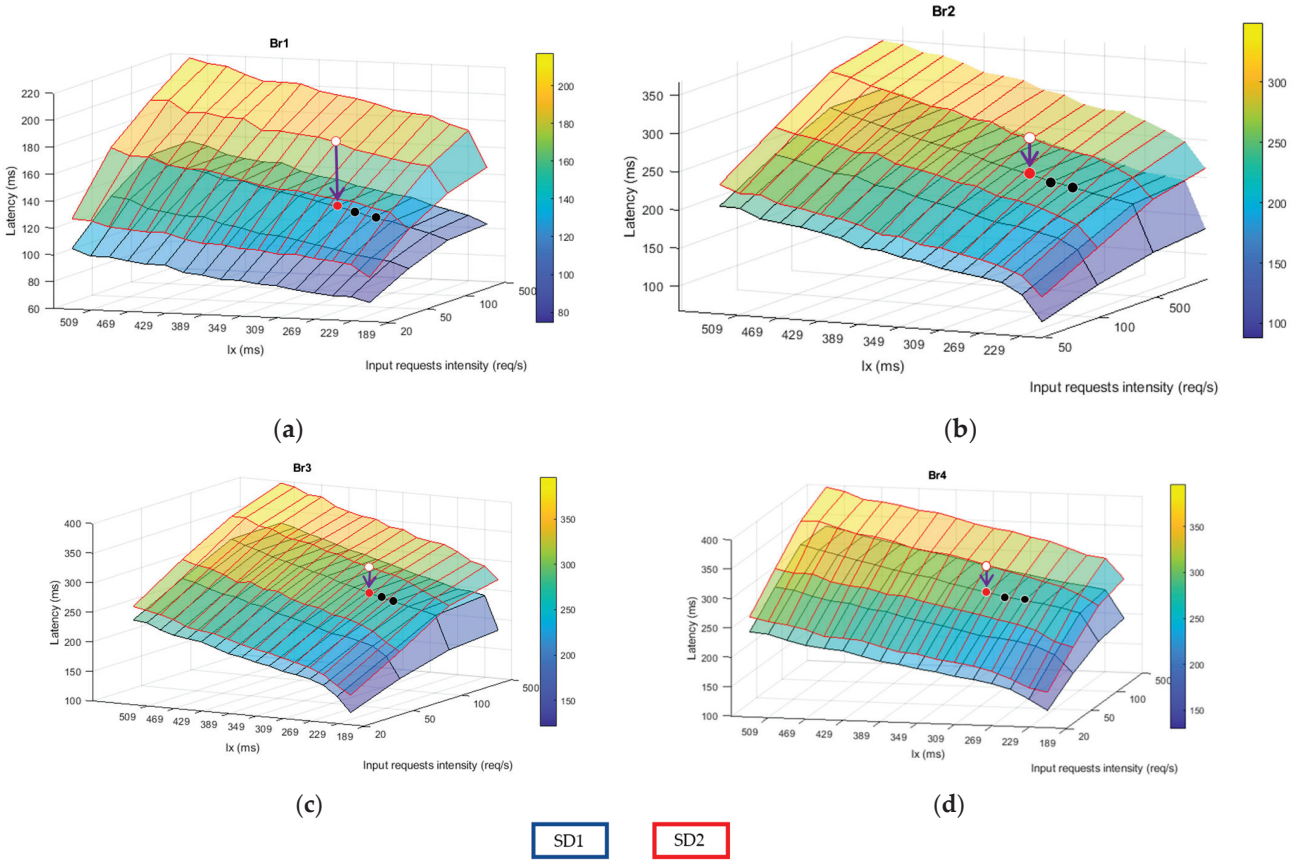


Figure 7. (**a**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_1 of MCS A. (**b**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_2 of MCS A. (**c**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_3 of MCS A. (**d**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_4 of MCS A.

Figures 7–9 show that within each MCS, the mean of $Latency_x(Ngl(to))$ can vary significantly with changes in the adjustment interval I, due to variations in the residual window ΔI_x, even when the intensity of the incoming request flow $UtoBr$ remains constant. Moreover, the magnitude of these changes is not uniform, even within a single MCS. For example, with SD1 and $UtoBr$ = 20 req/s:

- In MCS A, an increase of 188% in the adjustment interval I resulted in an increase in $Latency_x(Ngl(to))$ of 42% in Br_1, 136% in Br_2, 100% in Br_3, and 88% in Br_4;
- In MCS B, an increase of 134% in I led to an increase in $Latency_x(Ngl(to))$ of 14% in Br_1, 50% in Br_2, 65% in Br_3, and 84% in Br_4;
- In MCS C, an increase of 104% in I caused an increase in $Latency_x(Ngl(to))$ of 14% in Br_1, 31% in Br_2, 42% in Br_3, and 40% in Br_4.

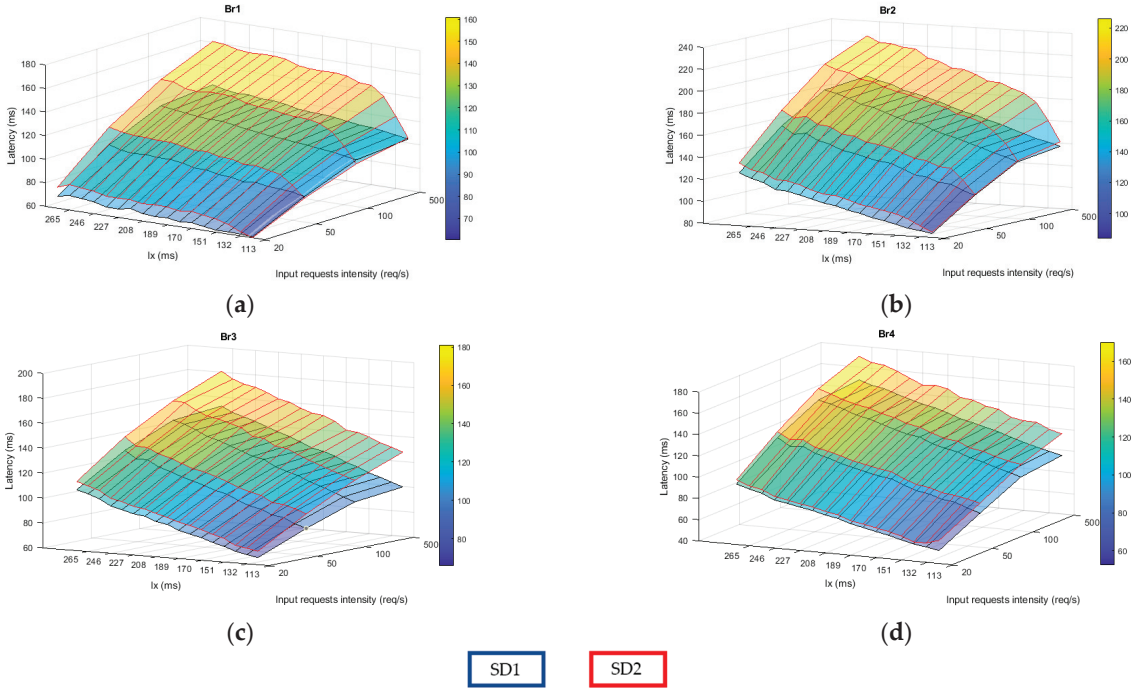


Figure 8. (**a**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_1 of MCS B. (**b**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_2 of MCS B. (**c**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_3 of MCS B. (**d**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_4 of MCS B.

The observed differences in the absolute values of $Latency_x(Ngl(to))$ among brokers within each MCS are explained by the mutual influence of the random duration of the packet transmission between brokers. Thus, the obtained results indicate that the varying interval I relative to the residual window ΔI_x is a significant factor in developing scenarios for adapting the data write latency.

The general conclusion from the results in Figures 7–9 is that there exists a positive correlation between the SD of $\Delta t^{tr}_{from-to}$ and the mean of the latency of the formation of the correct packet number. This implies that, for example, an increase in the SD of $\Delta t^{tr}_{from-to}$ at a fixed average of $\Delta t^{tr}_{from-to}$ leads to an increase in $Latency_x(Ngl(to))$. Therefore, when generating adaptation scenarios for data write latency, it becomes necessary to revise W_x, ΔI_x, and I. It should be noted that when changing the average of $\Delta t^{tr}_{from-to}$, new experiments must be conducted to estimate the average of latency of the formation of the correct packet numbers and the selection of the optimal interval I with the subsequent selection of the windows W_x and ΔI_x in each broker.

Finally, the comparison of the results in Figures 7–9 shows that for each intensity *UtoBr* of the incoming request flow, there exists a distinct rate and trajectory of influence of ΔI_x within the adjustment interval I on the $Latency_x(Ngl(to))$. This emphasizes the importance of forecasting changes in the parameters of the incoming request flow in order to design individualized scenarios for adapting data write latency in the given MCS.

4.3. Group of Experiments "varW"

In the experiments of the "varW" group, the dependence of the latency of forming the correct packet numbers on changes in the width of the availability window W_x within the adjustment interval I was investigated at different intensities *UtoBr*.

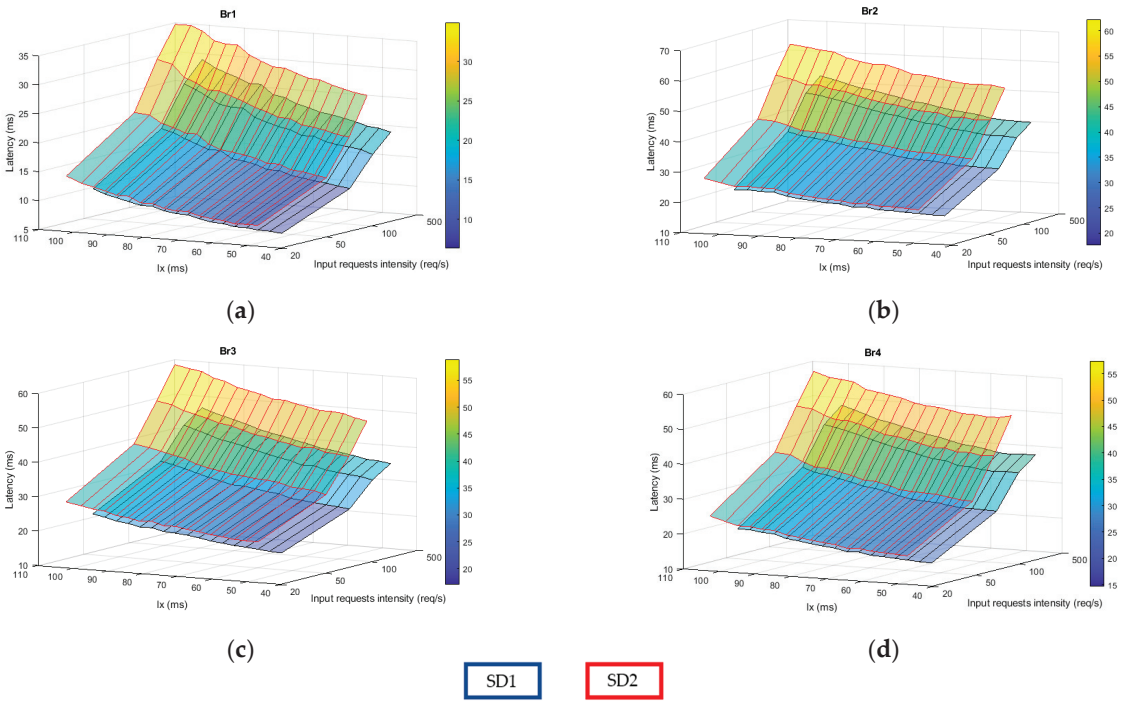


Figure 9. (**a**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_1 of MCS C. (**b**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_2 of MCS C. (**c**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_3 of MCS C. (**d**) Mean of the packet number formation latency (ms) over incoming request intensity (req/s) and adjustment interval I_x (ms) for the SD1 and SD2 sets in Br_4 of MCS C.

4.3.1. Setup of Experiments "varW" in the MCSs

In the experiments of the "varW" group, the values of W_x for the brokers of each MCS were organized into predefined sets. In this case, for each adjustment interval I_x, $x = 1 \ldots 7$ from the range, the boundary values (Table 4) and sets of W_x, named $\{W_0 - 50\%, W_0 - 25\%, W_0, W_0 + 25\%, W_0 + 50\%\}$ (Table 6), were used.

Table 6. Sets of availability windows W in MCS A, B, and C for experiments "varW".

MCS	Name of Set	Availability Windows W_x (ms)			
		W_1	W_2	W_3	W_4
A	$W_0 - 50\%$	84.9350	34.8850	10.8750	5.4700
	$W_0 - 25\%$	127.4025	52.3275	16.3125	8.2050
	W_0	169.8700	69.7700	21.7500	10.9400
	$W_0 + 25\%$	212.3375	87.2125	27.1875	13.6750
	$W_0 + 50\%$	254.8050	104.6550	32.6250	16.4100
B	$W_0 - 50\%$	17.5450	13.6700	5.5450	5.3650
	$W_0 - 25\%$	26.3175	20.5050	8.3175	8.0475
	W_0	35.0900	27.3400	11.0900	10.7300
	$W_0 + 25\%$	43.8625	34.1750	13.8625	13.4125
	$W_0 + 50\%$	52.6350	41.0100	16.6350	16.0950
C	$W_0 - 50\%$	10.0000	7.0000	6.0000	5.0000
	$W_0 - 25\%$	15.0000	10.5000	9.0000	7.5000
	W_0	20.0000	14.0000	12.0000	10.0000
	$W_0 + 25\%$	25.0000	17.5000	15.0000	12.5000
	$W_0 + 50\%$	30.0000	21.0000	18.0000	15.0000

In the experiment of the "varW" group, the average of $\Delta t^{tr}_{from-to}$ between brokers, as presented in Table 3, were used, while the SD of $\Delta t^{tr}_{from-to}$ was taken from the SD1 set from Table 5.

In each experiment, the input stream to each broker consisted of 10,000 user requests for data writing.

4.3.2. Results of Experiments "varW"

The results of the experiments of the "varW" group are presented in Figures 10–12 in the form of four surfaces. The upper surface in all cases was obtained at *UtoBr* = 500 req/s, and all surfaces below were obtained at {100, 50, 20} req/s, respectively.

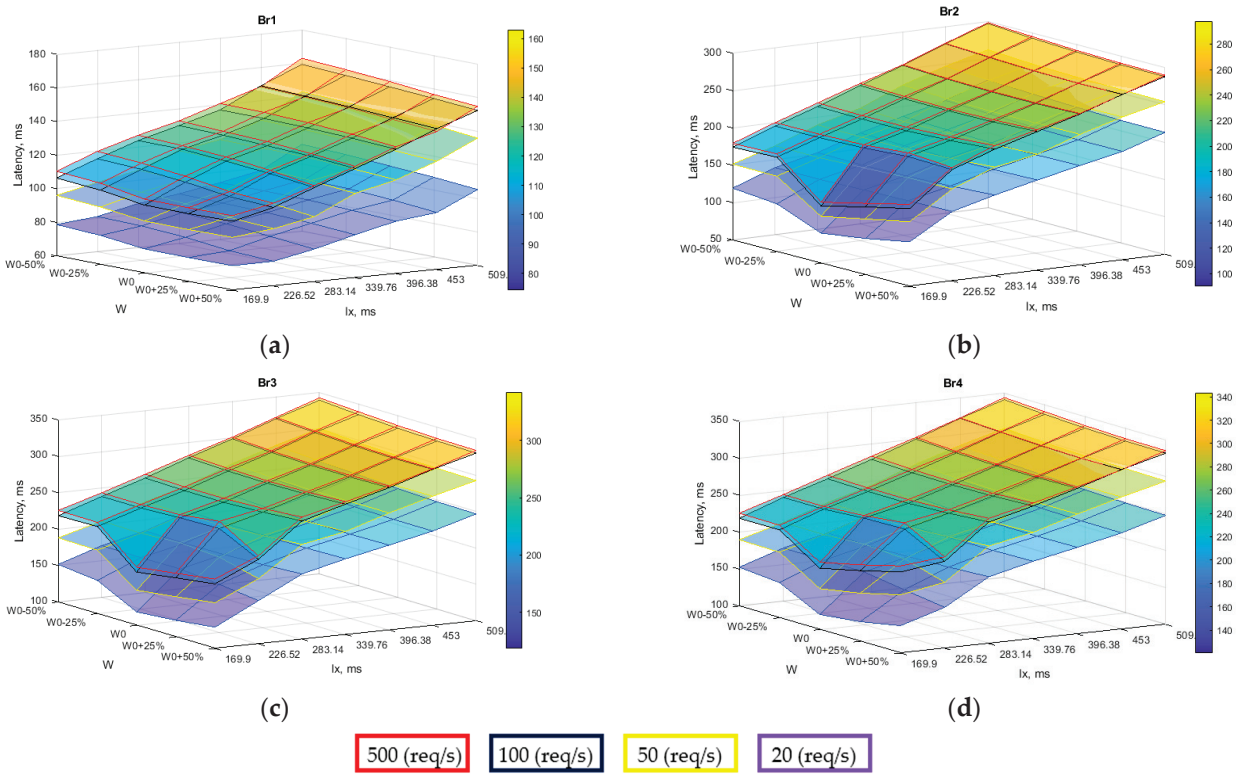


Figure 10. (**a**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_1 of MCS A. (**b**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_2 of MCS A. (**c**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_3 of MCS A. (**d**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_4 of MCS A.

The average of the latency of generating packet numbers $Latency_x(Ngl(to))$ for the base set W_0, as shown in Figures 10–12, is equal to the results of the experiments in Figures 7–9 for SD1 across all MCSs. The results in Figures 10–12 demonstrate that when wider availability windows are used for the same adjustment interval I, i.e., the W_0 + 50% sets, the latency $Latency_x(Ngl(to))$ in all MCSs is lower than with narrower windows, i.e., the W_0 − 50% sets. However, the magnitude of reduction in latency $Latency_x(Ngl(to))$ at availability window widening is not uniform for different durations of the adjustment interval I at the same request intensity *UtoBr*. For example, with intensity *UtoBr* = 500 req/s, expanding the availability window from W_0−50% to W_0+50% resulted in a decrease in $Latency_x(Ngl(to))$:

- In Br_1 by 5% at I = 489.900 ms and by 6% at I = 169.900 ms; in Br_4 by 2.5% at I = 489.900 ms and by 11% at I = 169.900 ms in MCS A;
- In Br_3 by 16% at I = 265.250 ms and by 9% at I = 113.250 ms; in Br_4 by 6% at I = 265.250 ms and by 5% at I = 113.250 ms in MCS B;
- In Br_1 by 16% at I = 108.525 ms and by 8% at I = 53.125 ms; in Br_4 by 10% at I = 108.525 ms and by 4% at I = 53.125 ms in MCS C.

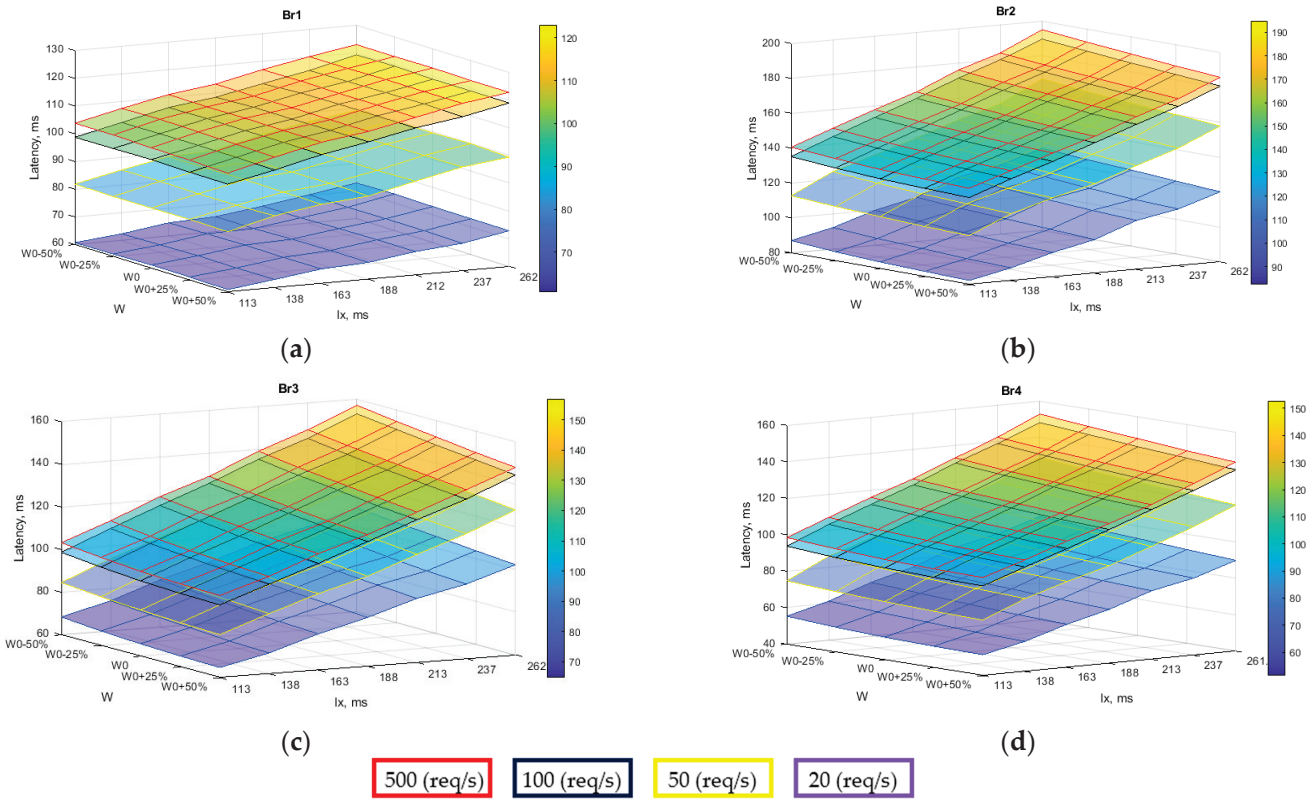


Figure 11. (**a**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_1 of MCS B. (**b**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_2 of MCS B. (**c**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_3 of MCS B. (**d**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_4 of MCS B.

Thus, modifying the width of the availability window W_x without altering the adjustment interval I can significantly influence $Latency_x(Ngl(to))$ and can therefore be employed in forming individual scenarios for adapting the data write latency in the MCSs.

The analysis of latency $Latency_x(Ngl(to))$ changes between the surfaces in Figures 10–12 for all MCSs shows that the increase in $Latency_x(Ngl(to))$ with increasing incoming request intensity *UtoBr* is not uniform. For example, in MCS C (Figure 12), an increase in *UtoBr* from 100 req/s to 500 req/s (i.e., fivefold) led to a maximal increase in $Latency_x(Ngl(to))$ of ≈16%. In contrast, in MCS A (Figure 10), an increase in *UtoBr* from 20 req/s to 50 req/s (i.e., 2.5-fold) produced a maximal increase in $Latency_x(Ngl(to))$ of 29%. The obtained results not only emphasize the individual nature of the influence of the parameters of the incoming request flow on the brokers but also demonstrate the necessity of accounting for the rate of change in *UtoBr* when creating scenarios for adapting the data write latency.

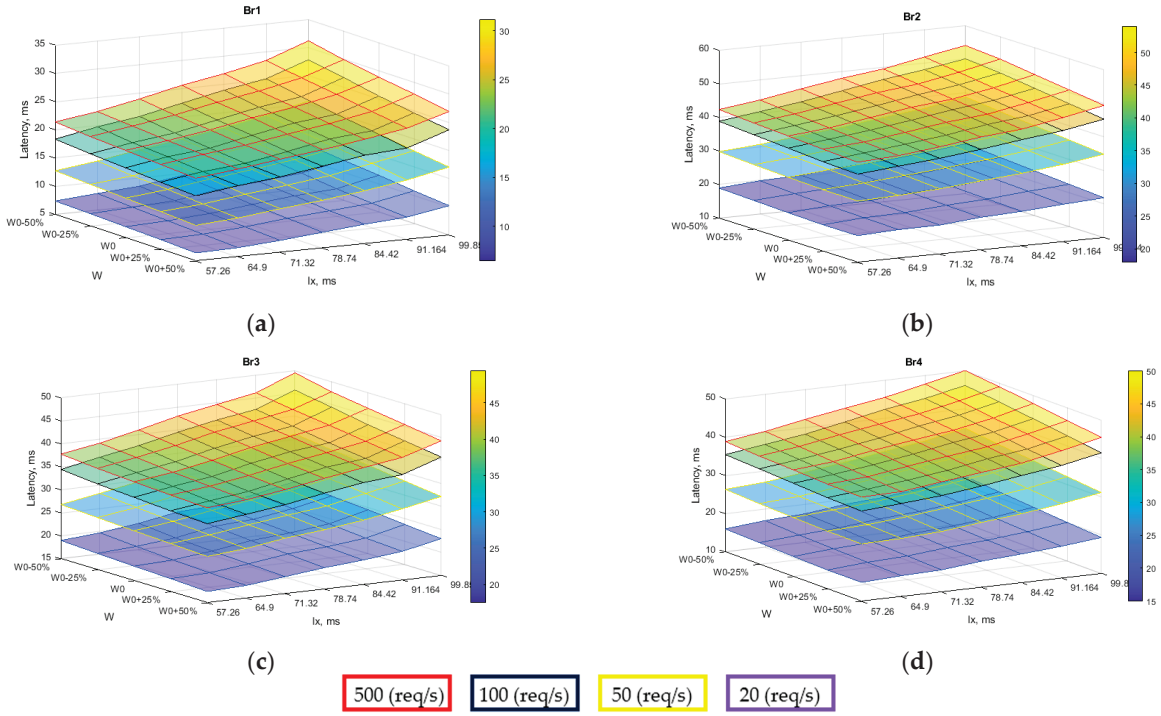


Figure 12. (**a**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_1 of MCS C. (**b**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_2 of MCS C. (**c**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_3 of MCS C. (**d**) Mean of the packet number formation latency (ms) over adjustment interval I_x (ms) and the sets of windows W for different incoming request intensities (req/s) in Br_4 of MCS C.

5. Discussion

The analysis of the results presented in the section above confirms the hypothesis about the presence of a significant influence of the internal parameters of the Latord method of data consistency on one distinct component of the data write latency—the latency $Latency_x(Ngl(to))$. In addition, the results demonstrated a non-uniform influence on the $Latency_x(Ngl(to))$ not only of the intensity *UtoBr* but also of the duration of packet transmission between brokers, which is another component of $LatencyWr(Ngl(to))$ from expression (9). Thus, the experimental results demonstrate that the choice of the values of the internal parameters of the Latord method that provide the required level of data write latency in replicas is a multi-criteria optimization problem.

In the conducted simulation experiments, it was assumed that the user request flows to different brokers follow the same probability distribution parameters within a single MCS. Under real conditions, each broker receives a non-stationary incoming request flow, and the data write latency needs to be adapted to the fluctuations in the incoming request flow. By analogy, fluctuations in user activity in real cases may not occur simultaneously at all brokers, but only at one or several of them. This demonstrates the need to develop individual scenarios for adapting the data write latency for each MCS based on forecasts of changes in external parameters and data on the influence of elements of the set $U(t)$ on the addenda of expression (9) separately.

Let us consider the principles of forming an adaptation scenario for the data write latency in MCS A based on the results of the experiments shown in Figure 7. Let the MCS operate with the following conditions: I = 289.90 ms, the SD of $\Delta t^{tr}_{from-to}$ corresponds to

the SD2 set from Table 5, and the intensity of the incoming request flow to each broker is 100 req/s. The average of the data write latency in each broker with such operating parameters in MCS A, calculated using Formula (5), is given in Table 7.

Table 7. Average of the data write latency in MCS A for I = 289.90 ms, SD2 of $\Delta t^{tr}_{from-to}$, and *UtoBr* = 100 req/s.

Average of the Data Write Latency (ms)			
Br_1	Br_2	Br_3	Br_4
664.6392	682.7719	713.6057	724.7734

The average values of the duration of the formation of the correct packet number in each broker of MCS A for this configuration are presented in Table 8 and are shown as white points on the upper surfaces in Figure 7.

Table 8. Average of the latency of the packet number formation in MCS A brokers for different SD values of packet transmission durations between brokers.

Adjustment Interval I (ms)	Standard Deviation (ms)	Average of the Latency of the Formation of Packets Number (ms)			
		Br_1	Br_2	Br_3	Br_4
289.90	SD2	144.1962	227.7739	259.4436	261.4491
	SD1	107.4606	189.0890	225.3060	227.1781
209.90	SD1	96.5219	167.9406	203.1084	205.3815
229.90		98.5903	173.5104	209.5391	212.5863
249.90		101.8672	179.1206	215.1092	216.7648
269.90		104.5551	182.0716	217.8441	219.7127

Suppose that the implementation of new technological solutions for dispersion compensation on one segment of the global communication line used for packet transmission between brokers in MCS A resulted in a reduction in the SD of $\Delta t^{tr}_{from-to}$, such that it now corresponds to the SD1 set from Table 5. All other characteristics of the external variables remained unchanged.

According to the results of the conducted experiments, a decrease in SD will lead to a decrease in the average latency of the formation of the packet number (Table 8) in each broker, resulting in values indicated by the red points on the lower surfaces in Figure 7. Consequently, this will require a secondary selection of the parameters of the Latord method, namely, I, W_x, and ΔI_x. The selection of a new adjustment interval I from the neighboring candidate values, marked by black points in Figure 7, should be based not only on the mean of the latencies of the packet number formation but also on the corresponding estimates of the ballast interval τ_x in each broker. Choosing the interval I, for which the mean value of τ_x is close to zero, may increase the likelihood of packet loss during transmission between brokers. Table 9 presents the data write latency calculated using Formula (5) based on the conditions of the "varΔI" experiment group in MCS A for SD1 and for the adjustment intervals, among which a new value of the interval I may be selected.

The analysis of the results presented in Table 9 shows that the mean data write latency in each broker decreases as I decreases. However, at I < 229.90 ms, the number of packets received by passive brokers from the active one after *MaxLate* is more than 3.7% of the total number of sent packets, which means that the new value of I should be equal to 249.90 ms or 269.90 ms.

It should be noted that a decrease in the data write latency in brokers due to decreasing I, considered in this adaptation scenario, is possible if, before changing the external

operational parameters of the MCS, the chosen interval I value was the one that allow to its decrease, rather than the minimum acceptable interval I. This means that the decision regarding the Latord method's parameters that affect the data write latency must be made by taking into account the forecasts of their changes.

Table 9. Average of the data write latency in MCS A for different adjustment intervals.

Adjustment Interval I (ms)	Average of the Data Write Latency (ms)				Number of "Late" Packets out of 10,000
	Br_1	Br_2	Br_3	Br_4	
209.90	502.7898	487.6762	515.2307	523.9986	6696
229.90	539.7138	535.8083	564.8082	574.1685	1838
249.90	579.9570	584.8174	615.1671	624.2518	0
269.90	621.2052	632.7332	664.1750	673.3707	0
289.90	664.6258	682.4046	716.1017	724.1544	0

As can be seen from Table 9, at I = 249.90 ms, the average of the data write latency for all brokers is not less than that at I = 269.90 ms. To achieve a more acceptable data write latency, the possibilities of changing W_x within these two values of I can be assessed in the next step of the adaptation scenario. The experimental results shown in Figure 10 suggest that adjusting the ratio between windows W_x and ΔI_x within these intervals I is expected to further reduce the data write latency in MCS A. The average values of the data write latency for such adjustment intervals I under different availability window sets are presented in Table 10.

Table 10. Average of the data write latency in MCS A for different adjustment intervals I and availability window sets W.

Adjustment Interval I (ms)	Name of Set	Average of the Data Write Latency (ms)			
		Br_1	Br_2	Br_3	Br_4
249.90	$W_0 - 50\%$	578.1322	604.7697	627.1246	627.8015
	$W_0 - 25\%$	**571.9400**	594.0697	621.5324	625.6943
	W_0	579.9570	584.8174	615.1671	624.2518
	$W_0 + 25\%$	601.1039	**577.7559**	**611.6084**	**620.2109**
	$W_0 + 50\%$	637.6934	573.0511	606.0604	617.2175
269.90	$W_0 - 50\%$	627.0558	655.2516	675.6664	679.8482
	$W_0 - 25\%$	618.0937	643.8742	670.4398	677.5652
	W_0	621.2052	632.7332	664.1750	673.3707
	$W_0 + 25\%$	635.8850	626.6528	660.3239	671.7383
	$W_0 + 50\%$	666.5338	620.0096	656.6366	669.9980

The analysis of the data in Table 10 shows that using the set window $W_0 - 25\%$ instead of the set W_0 for the adjustment interval I = 249.90 ms reduces the mean of the data write latency only in broker Br_1, which is highlighted in bold in Table 10, while in brokers Br_2–Br_4, it leads to an increase. A similar situation occurs when using the $W_0 + 25\%$ instead of W_0: the mean of the data write latency decreases in brokers Br_2–Br_4, which is highlighted in bold in Table 10, but increases in broker Br_1. In this case, the selection of the adjustment interval I and the corresponding set of windows can be based on the vector $[Signif_x]$ of the brokers' significance. If reducing the average of the data write latency is more important for users from the cluster with broker Br_1 than for other users, then the $W_0 - 25\%$ set of the availability windows at I = 249.90 ms should be selected. If any of the Br_2–Br_4 brokers is more important, then the $W_0 + 25\%$ set at I = 249.90 ms would be more optimal.

The considered adaptation scenario will reduce the data write latency when changing the SD of $\Delta t^{tr}_{from-to}$ in MCS A by ≈13% in all brokers when using the $W_0 - 25\%$ set of availability windows or by ≈9% in broker Br_1 and by 14–15% in brokers Br_2–Br_4 when using the set $W_0 + 25\%$.

The considered adaptation scenario also emphasizes the individual nature of changes in the data write latency when the external conditions of the MCS operation fluctuate.

It should be noted, however, that in the simulation experiments, it was assumed that the packet transmission times between brokers follow the same probability distribution with a common SD value for all brokers. Moreover, the SD was varied simultaneously and to the same level in all brokers at once. Under real conditions, changes in the average and SD values may occur for one or more broker pairs and by different amounts, which should be taken into account in the predictive AI model.

In conclusion, the Latord method of data consistency demonstrates robust capabilities for generating the adaptation scenarios of the data write latency in geo-distributed MCSs, even when the intensity of incoming request flow changes by 50 times (Figures 7–9). This property is particularly valuable for load scaling, making the proposed approach a promising solution for adaptive latency management in MCSs.

6. Conclusions and Future Work

We present an approach to creating scenarios for adapting the data write latency in geo-distributed MCSs using learning AI models to solve a multi-criteria optimization problem. The ability to control data write latency is achieved by selecting optimal parameters of the Latord method of data consistency in the presence of data on changes in incoming packet flows and in the parameters of packet transmission between brokers. Simulation experiments in three MCSs of different configurations demonstrate the need to use individual scenarios for adapting the data write latency; for its formation, a scheme of interaction between three different learning AI models is proposed.

Future work will focus on studying the impact of non-uniform external influences on the data write latency in MCSs using the algorithmic data consistency method. Non-uniform changes in the incoming user requests more closely reflect the real operating conditions of the MCSs. This will allow us to determine the structure and parameters of the AI model for determining the minimum gradients from the set $Y(t)$ and, therefore, create more effective scenarios for adapting the data write latency in replicas to changing external conditions.

Author Contributions: Conceptualization, O.K. and M.K.; methodology and software, M.K.; validation, O.K. and M.V.; formal analysis, M.I.; resources, J.M.; writing—original draft preparation, O.K.; writing—review and editing, V.P.; visualization, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are contained within the article.

Conflicts of Interest: Author Olha Kozina was employed by the AFOREHAND Studio. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Saili Krishna, M. Multi-Cloud Automation: A Strategic Approach to Cloud Infrastructure Management. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2024**, *10*, 183–190. [CrossRef]
2. Hong, J.; Dreibholz, T.; Schenkel, J.A.; Hu, J.A. An Overview of Multi-Cloud Computing. In *Web, Artificial Intelligence and Network Applications*; Advances in Intelligent Systems and, Computing; Barolli, L., Takizawa, M., Xhafa, F., Enokido, T., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 927, pp. 1055–1068.
3. Rafique, A.; Joosen, W.; Lagaisse, B. Middleware for Data Management in Multi-Cloud. Ph.D. Thesis, KU Leuven, Leuven, Belgium, 2019; p. 222. Available online: https://lirias.kuleuven.be/retrieve/531841 (accessed on 19 August 2025).
4. Campêlo, R.A.; Casanova, M.A.; Guedes, D.O.; Laender, A.H.F. A Brief Survey on Replica Consistency in Cloud Environments. *J. Internet Serv. Appl.* **2020**, *11*, 1–13. [CrossRef]
5. Mokadem, R.; Hameurlain, A. A Data Replication Strategy with Tenant Performance and Provider Economic Profit Guarantees in Cloud Data Centers. *J. Syst. Softw.* **2020**, *159*, 110447. [CrossRef]
6. Fazlina, M.A.; Latip, R.; Ibrahim, H.; Abdullah, A. Replication Strategy with Comprehensive Data Center Selection Method In Cloud Environments. *Comput. Mater. Contin.* **2023**, *74*, 415–433.
7. Shithil, S.M.; Adnan, M.A. A Prediction Based Replica Selection Strategy for Reducing Tail Latency in Geo-Distributed Systems. *IEEE Trans. Cloud Comput.* **2023**, *11*, 2954–2965. [CrossRef]
8. Najjar, A.; Mokadem, R.; Pierson, J.-M. A Review of Data Placement and Replication Strategies Based on Machine Learning. In Proceedings of the 2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS), Belgrade, Serbia, 10–14 October 2024; IEEE: New York, NY, USA, 2024; pp. 278–285.
9. Wang, H.; Shen, H.; Li, Z.; Tian, S. GeoCol: A Geo-Distributed Cloud Storage System with Low Cost and Latency Using Reinforcement Learning. In Proceedings of the 2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS), Washington, DC, USA, 7–10 July 2021; IEEE: New York, NY, USA, 2021; pp. 149–159.
10. Mokadem, R.; Gil, J.M.; Hameurlain, A.; Kueng, J. A Review on Data Replication Strategies in Cloud Systems. *IJGUC* **2022**, *13*, 347. [CrossRef]
11. Tos, U.; Mokadem, R.; Hameurlain, A.; Ayav, T.; Bora, S. A Performance and Profit Oriented Data Replication Strategy for Cloud Systems. In Proceedings of the 2016 Intnational IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), Toulouse, France, 18–21 July 2016; IEEE: New York, NY, USA, 2016; pp. 780–787.
12. Bernardin, S.H.; Mokadem, R.; Morvan, F.; Ramanana, H.; Rakotoarivelo, H. TCDRM: A Tenant Budget-Aware Data Replication Framework for Multi-Cloud Computing. *JLISS* **2025**, *12*, 3.
13. Azure Network Round-Trip Latency Statistics. Available online: https://learn.microsoft.com/en-us/azure/networking/azure-network-latency (accessed on 8 July 2025).
14. AWS Latency Monitoring. Available online: https://www.cloudping.co/ (accessed on 8 July 2025).
15. Wu, Z.; Madhyastha, H.V. Understanding the Latency Benefits of Multi-Cloud Webservice Deployments. *SIGCOMM Comput. Commun. Rev.* **2013**, *43*, 13–20. [CrossRef]
16. Eischer, M.; Straßner, B.; Distler, T. Low-Latency Geo-Replicated State Machines with Guaranteed Writes. In Proceedings of the 7th Workshop on Principles and Practice of Consistency for Distributed Data, Heraklion, Greece, 27 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–9.
17. Coelho, P.; Pedone, F. GeoPaxos+: Practical Geographical State Machine Replication. In Proceedings of the 2021 40th International Symposium on Reliable Distributed Systems (SRDS), Chicago, IL, USA, 20–23 September 2021; IEEE: New York, NY, USA, 2021; pp. 233–243.
18. Whittaker, M.; Charapko, A.; Hellerstein, J.M.; Howard, H.; Stoica, I. Read-Write Quorum Systems Made Practical. In Proceedings of the 8th Workshop on Principles and Practice of Consistency for Distributed Data, Online, 26 April 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1–8.
19. Charapko, A.; Ailijiang, A.; Demirbas, M. PigPaxos: Devouring the Communication Bottlenecks in Distributed Consensus. In Proceedings of the 2021 International Conference on Management of Data, Virtual, 20–25 June 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 235–247.
20. Song, H.; Wang, Y.; Chen, X.; Feng, H.; Feng, Y.; Fang, X.; Cui, H.; Kong, L. K2: On Optimizing Distributed Transactions in a Multi-Region Data Store with TrueTime Clocks (Extended Version). *arXiv* **2025**, arXiv:2504.01460. [CrossRef]
21. Lu, H.; Mu, S.; Sen, S.; Lloyd, W. NCC: Natural Concurrency Control for Strictly Serializable Datastores by Avoiding the Timestamp-Inversion Pitfall. *arXiv* **2023**, arXiv:2305.14270. [CrossRef]
22. Geng, J.; Sivaraman, A.; Prabhakar, B.; Rosenblum, M. Nezha: Deployable and High-Performance Consensus Using Synchronized Clocks. *Proc. VLDB Endow.* **2022**, *16*, 629–642. [CrossRef]

23. Kozina, O.A.; Panchenko, V.I.; Kolomiitsev, O.V.; Usik, V.V.; Stratiienko, N.K.; Safoshkina, L.V.; Kucherenko, Y.F. Data Consistency Protocol for Multicloud Systems. *IJCC* **2024**, *13*, 42–61. [CrossRef]
24. Volk, M.; Kozina, O.; Buhrii, A.; Osiievskyi, S.; Kozin, M.; Volk, D.; Diachenko, D.; Turinskyi, Y. Devising a Method for Data Consistency at Replication in Multicloud Systems. *East.-Eur. J. Enterp. Technol.* **2025**, *4*, 14–22.
25. Seth, D.K.; Ratra, K.K.; Sundareswaran, A.P. AI and Generative AI-Driven Automation for Multi-Cloud and Hybrid Cloud Architectures: Enhancing Security, Performance, and Operational Efficiency. In Proceedings of the 2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 6–8 January 2025; IEEE: New York, NY, USA, 2025; pp. 784–793.
26. Rossi, A.; Visentin, A.; Carraro, D.; Prestwich, S.; Brown, K.N. Forecasting Workload in Cloud Computing: Towards Uncertainty-Aware Predictions and Transfer Learning. *Clust. Comput.* **2025**, *28*, 258. [CrossRef]
27. Karpagam, T.; Kanniappan, J. Symmetry-Aware Multi-Dimensional Attention Spiking Neural Network with Optimization Techniques for Accurate Workload and Resource Time Series Prediction in Cloud Computing Systems. *Symmetry* **2025**, *17*, 383. [CrossRef]
28. Xu, P.; Goteng, G.L.; He, Y. Modelling Cloud Service Latency and Availability Using a Deep Learning Strategy. *Expert. Syst. Appl.* **2021**, *182*, 115121. [CrossRef]
29. Volk, M.; Kozina, O.; Kozin, M. Method for Synchronizing Data Write Requests in Federated Cloud Systems. *IIM* **2025**, *1*, 80–98. [CrossRef]
30. Kleppmann, M. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*; O'Reilly Media: Sebastopol, CA, USA, 2017.
31. Ibrahim, A.A.Z.A.; Kliazovich, D.; Bouvry, P. Service Level Agreement Assurance between Cloud Services Providers and Cloud Customers. In Proceedings of the2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Cartagena, Colombia, 16–19 May 2016; IEEE: New York, NY, USA, 2016; pp. 588–591.
32. Chana, I.; Singh, S. Quality of Service and Service Level Agreements for Cloud Environments: Issues and Challenges. In *Cloud Computing*; Computer Communications and, Networks; Mahmood, Z., Ed.; Springer International Publishing: Cham, Switzerland, 2014; pp. 51–72.
33. Mechouche, J.; Touihri, R.; Sellami, M.; Gaaloul, W. Conformance Checking for Autonomous Multi-Cloud SLA Management and Adaptation. *J. Supercomput.* **2022**, *78*, 13004–13039. [CrossRef]
34. Son, S.; Choi, H.-H.; Oh, B.T.; Kim, S.W.; Kim, B.S. Cloud SLA Relationships in Multi-Cloud Environment: Models and Practices. In Proceedings of the 8th International Conference on Computer Modeling and Simulation, Canberra, Australia, 20–23 January 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1–6.
35. Sakurada, L.; De La Prieta, F.; Leitao, P. The Role of Multi-Agent Systems in Realizing Asset Administration Shell Type 3. *Future Internet* **2025**, *17*, 270. [CrossRef]
36. Martínez, G.; Hernández, J.A.; Reviriego, P.; Reinheimer, P. Round Trip Time (RTT) Delay in the Internet: Analysis and Trends. *IEEE Netw.* **2023**, *38*, 280–285. [CrossRef]
37. Coimbra, M.E.; Selimi, M.; Francisco, A.P.; Freitag, F.; Veiga, L. Gelly-Scheduling: Distributed Graph Processing for Service Placement in Community Networks. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Pau, France, 9–13 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 151–160.
38. Landa, R.; Clegg, R.G.; Araujo, J.T.; Mykoniati, E.; Griffin, D.; Rio, M. Measuring the Relationships between Internet Geography and RTT. In Proceedings of the 2013 22nd International Conference on Computer Communication and Networks (ICCCN), Nassau, The Bahamas, 30 July–2 August 2013; IEEE: New York, NY, USA, 2013; pp. 1–7.
39. Newman, M.E.J. Power Laws, Pareto Distributions and Zipf's Law. *Contemp. Phys.* **2005**, *46*, 323–351. [CrossRef]
40. Gandica, Y.; Carvalho, J.; Sampaio Dos Aidos, F.; Lambiotte, R.; Carletti, T. Stationarity of the Inter-Event Power-Law Distributions. *PLoS ONE* **2017**, *12*, e0174509. [CrossRef] [PubMed]
41. Downey, A.B. Lognormal and Pareto Distributions in the Internet. *Comput. Commun.* **2005**, *28*, 790–801. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

Deep Reinforcement Learning for Adaptive Robotic Grasping and Post-Grasp Manipulation in Simulated Dynamic Environments

Henrique C. Ferreira [1] and Ramiro S. Barbosa [1,2,*]

[1] Department of Electrical Engineering, Institute of Engineering—Polytechnic of Porto (ISEP/IPP), 4249-015 Porto, Portugal; 1180528@isep.ipp.pt
[2] GECAD—Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, ISEP/IPP, 4249-015 Porto, Portugal
* Correspondence: rsb@isep.ipp.pt

Abstract

This article presents a deep reinforcement learning (DRL) approach for adaptive robotic grasping in dynamic environments. We developed UR5GraspingEnv, a PyBullet-based simulation environment integrated with OpenAI Gym, to train a UR5 robotic arm with a Robotiq 2F-85 gripper. Soft Actor-Critic (SAC) and Proximal Policy Optimization (PPO) were implemented to learn robust grasping policies for randomly positioned objects. A tailored reward function, combining distance penalties, grasp, and pose rewards, optimizes grasping and post-grasping tasks, enhanced by domain randomization. SAC achieves an 87% grasp success rate and 75% post-grasp success, outperforming PPO 82% and 68%, with stable convergence over 100,000 timesteps. The system addresses post-grasping manipulation and sim-to-real transfer challenges, advancing industrial and assistive applications. Results demonstrate the feasibility of learning stable and goal-driven policies for single-arm robotic manipulation using minimal supervision. Both PPO and SAC yield competitive performance, with SAC exhibiting superior adaptability in cluttered or edge cases. These findings suggest that DRL, when carefully designed and monitored, can support scalable learning in manipulation tasks.

Keywords: robotic grasping; reinforcement learning; deep reinforcement learning (DRL); adaptive manipulation; robotic simulation; proximal policy optimization (PPO); soft actor-critic (SAC); deep deterministic policy gradient (DDPG); autonomous robotics; physics-based simulation

1. Introduction

Robotic grasping refers to the autonomous ability of a robotic manipulator to detect, plan, and execute actions that enable secure physical interaction with objects of various geometries and materials. In unstructured environments, this task becomes significantly more challenging due to the unpredictability of object placement, occlusions, dynamic disturbances, and variability in object properties. A successful grasp must also consider stability over time, task relevance (e.g. correct pose for assembly), and minimal risk of collision or slippage [1].

Robotic grasping in unstructured environments is essential for industrial automation, assistive robotics, and human–robot interaction [2]. Applications such as warehouse sorting, assistive devices, and assembly require robust grasping of diverse objects (e.g., 0.02–0.05 m

cubes, cylinders, friction [0.5,1.2] under dynamic conditions. Analytical methods such as Ferrari and Canny [3] achieve 50–60% success for novel objects but fail in cluttered scenes (35% failure under occlusions, 40% under perturbations) [4]. DRL enables adaptive policies, handling variability in shape, texture, and dynamics [5,6]. However, the Levine et al. DRL system [7] achieves 80% success, but requires 800,000 grasps and 10,000 GPU hours, limiting scalability. These methods neglect post-grasping tasks such as reorientation, critical for assembly [2,8]. The sim-to-real gap, due to PyBullet idealized physics [9], reduces the transfer success to 65–70% [10].

Recent research has produced advances in deep learning-based perception and control, enabling robots to generalize across unseen instances. However, existing approaches either rely heavily on supervision, operate in constrained settings, or require expensive data collection and compute infrastructure. In particular, end-to-end DRL systems lack integrated handling of post-grasping behaviors, and their transferability to real robots remains limited due to the sim-to-real gap.

This work proposes a DRL-based architecture that enables incremental learning of grasping and post-placement behaviors in a simulated environment, specifically designed to encourage object diversity and realistic interaction constraints. Our approach builds on state-of-the-art actor–critic algorithms and leverages domain randomization to enhance transferability. It presents a DRL-based framework using a UR5 arm with a Robotiq 2F-85 gripper in UR5GraspingEnv. The objectives include (i) developing a scalable simulation, (ii) comparing PPO [11] and SAC [12], and (iii) addressing sim-to-real via domain randomization [13]. The evaluation of grasp stability metrics is also discussed. Although UR5GraspingEnv mirrors the interfaces of a physical UR5 via ROS, this study does not report hardware tests. Therefore, deployment claims are presented as potential for transfer, subject to future validation on the real platform under a defined calibration and safety protocol. Table 1 lists the application requirements.

The main contributions of this work are as follows:

- We design and validate a modular DRL simulation environment tailored to real-world grasping and post-grasping challenges, using PyBullet and OpenAI Gym.
- We implement and benchmark two state-of-the-art algorithms, SAC and PPO, highlighting their respective advantages in terms of policy stability, adaptability, and convergence.
- We introduce task-specific reward shaping and domain randomization techniques to enhance policy generalization and improve transfer potential.
- We provide a structured evaluation methodology based on grasp stability, object positioning, and task success, with results that inform future deployment.

Table 1. Application of requirements for robotic grasping applications.

Application	Success Rate (%)	Challenges
Warehouse Sorting	90	Dynamic objects, clutter
Assistive Robotics	85	Variable textures, safety
Assembly Lines	95	Post-grasp reorientation

Our work introduces several key contributions that distinguish it from previous research. Unlike studies such as Levine et al. [7] and QT-Opt (Kalashnikov et al. [14]), which focused on single algorithmic approaches, we provide a unified comparative evaluation of two widely used DRL algorithms, PPO and SAC, under identical conditions, allowing for a thorough analysis of success rates, convergence speed, stability, and robustness. Furthermore, unlike Dex-Net (Mahler et al. [15]) and James et al. [10], which primarily assess grasp success, we integrate post-grasp placement metrics to better reflect realistic industrial

scenarios involving sorting, assembly, and warehouse automation. We also extend domain randomization techniques, specifically targeting both grasping and placement tasks, thereby enhancing the robustness of learned policies under dynamic variations. Finally, we introduce UR5GraspingEnv, a modular and reproducible simulation benchmark designed for extensibility, bridging research experimentation with practical deployment, and offering a valuable tool for the robotics community.

In this study, we intentionally restrict the baseline comparison to PPO and SAC, two of the most established DRL algorithms for continuous robotic control. This choice enables a deeper analysis of convergence behavior, stability, and generalization under identical conditions, providing clearer insights into the trade-offs between on-policy and off-policy methods. Although modern alternatives such as TD3, DreamerV3, or imitation learning are promising, they fall outside the present scope and are identified as directions for future work.

Furthermore, our work expands on James et al. [10] by explicitly addressing post-grasp placement performance and differs from Rajeswaran et al. [16] by combining domain randomization with placement-focused metrics. These extensions ensure that our contribution goes beyond grasp robustness to encompass full manipulation sequences, situating it as a holistic integration of algorithmic evaluation, robustness strategies, and environment design.

In addition to RGB-D integration and physical validation, the next stage will focus on extending the framework towards multi-object manipulation. By incorporating hierarchical rewards and curriculum learning, we aim to address the current bottleneck of reward shaping and enable scalable policies for complex assembly pipelines.

The remainder of this paper is organized as follows. Section 2 reviews related work on robotic grasping and DRL techniques. Section 3 details the simulation environment, system architecture, and learning algorithms. Section 4 presents the experimental setup, metrics, and results. Section 5 analyses the results and limitations. Finally, Section 6 concludes the article and outlines the directions for future work.

2. Related Work

Robotic grasp spans analytical, data-driven, and learning-based methods [2]. Analytical approaches such as Ferrari and Canny [3] achieve 50–60% success but fail in dynamic settings (40% failure) [4]. Deep learning methods improve adaptability [7,15]. Levine et al. [7] achieve 80% success but require 800,000 grasps. Mahler et al. Dex-Net 2.0 [15] reaches 90% success with synthetic data but neglects post-grasping. Akkaya et al. [8] solve dexterous tasks (85% success) with costly real-world training. Andrychowicz et al. [17] achieve 85% manipulation success with 5000 GPU hours.

DRL frameworks, pioneered by Mnih et al. [6], enable complex learning [18]. Lillicrap et al. DDPG [19] laid the foundations for continuous control, influencing SAC [12]. Kalashnikov et al. [14] use RGB-D to grasp with 85% success and 20% failure in cluttered scenes. James et al. [10] achieve 85% real-world success, dropping to 70% for dynamic objects. Pinto and Gupta [20] achieve 80% success with 50,000 tries, lacking post-grasping focus. Rajeswaran et al. [16] achieve 80% multi-object success with complex reward shaping. Table 2 compares methods.

Beyond these individual contributions, several recent surveys have systematized the progress in DRL-based manipulation and provided a broader context for our work. Han et al. [21] review over a decade of DRL methods in robotic manipulation, with particular emphasis on reward engineering, sample efficiency, and sim-to-real transfer. Their analysis highlights that while algorithms such as PPO and SAC dominate current benchmarks, the main bottlenecks remain in designing informative rewards and ensuring robustness un-

der uncertainty, two aspects directly addressed in our framework through multithreshold reward shaping and domain randomization. Song et al. [22] provide an in-depth review of learning-based dexterous grasping, categorizing approaches into grasp generation, execution, and evaluation. They stress that most works optimize grasp stability or dexterity in isolation, while neglecting post-grasp placement, which we explicitly integrate into our evaluation pipeline. Tang et al. [23] survey real-world deployments of DRL in robotics, identifying the gap between simulation-based successes and scalable industrial adoption. Their findings reinforce the importance of modular and reproducible environments that can bridge research prototypes and industrial deployment—a role directly fulfilled by the proposed UR5GraspingEnv.

Table 2. Comparison of prior robotic grasping methods.

Method	Success Rate (%)	Data/Compute	Limitations
Levine et al. [7]	80	800,000 grasps	Scalability, no post-grasp
Dex-Net 2.0 [15]	90	Synthetic data	Static focus
Akkaya et al. [8]	85	Real-world training	Compute cost
Kalashnikov et al. [14]	85	RGB-D data	Cluttered scenes
James et al. [10]	85	Simulation	Sim-to-real gap
Pinto and Gupta [20]	80	50,000 tries	No post-grasping
Rajeswaran et al. [16]	80	Complex rewards	Reward tuning

In particular, we can discuss Odeyemi et al. [24], who compare SAC, DQN, and PPO for intelligent prosthetic-hand grasping, and clarify how our work differs: (i) industrial single-arm manipulation with explicit post-grasp placement metrics, (ii) a unified, controlled comparison of PPO vs. SAC under identical conditions with domain randomization targeting both grasp and placement, and (iii) a reproducible UR5 benchmark (UR5GraspingEnv) designed for extensibility.

By situating our contribution within the context of these surveys, we emphasize how our work complements prior studies: we provide a reproducible benchmark for PPO and SAC, extend evaluation beyond grasp success to post-grasp manipulation, and demonstrate how domain randomization contributes to robustness. This positions our study not only as an incremental improvement, but as a framework that aligns with the open challenges identified in the most recent reviews.

Despite the success of previous work, most methods either (i) focus exclusively on grasp execution, neglecting post-grasp actions such as reorientation or placement, or (ii) demand vast computational resources and datasets for real-world training. Moreover, several approaches lack robustness under occlusion, clutter, and dynamic conditions.

This article addresses these limitations by developing a lightweight DRL framework in simulation, with support for post-grasping behaviors and scalable training using domain randomization. Our system, UR5GraspingEnv, enables benchmarking under controlled variability and serves as a bridge towards real-world deployment.

Specifically, we compare two popular off-policy and on-policy DRL algorithms—SAC [12] and PPO [11]—within the same manipulation environment, and evaluate their effectiveness in grasp success, stability, and object placement. By focusing on adaptability and reusability, this work complements the existing literature and addresses key open challenges in learning-based robotic grasping.

3. Methodology

This section details the DRL-based grasping system in UR5GraspingEnv.

3.1. Simulation Environment

UR5GraspingEnv, built with PyBullet [9] and OpenAI Gym [1], models a 6-DoF UR5 arm and a Robotiq 2F-85 gripper on a 1 m × 1 m × 0.5 m workspace. Objects (cubes, cylinders, spheres) have randomized sizes (0.02–0.05 m), masses (0.1–0.5 kg), and friction (0.3–0.7). Domain randomization [13] improves transferability.

The ranges selected for the objects were carefully chosen. Object dimensions (edge lengths 4–10 cm, masses 50–300 g) reflect the distribution of lightweight industrial parts and household objects commonly manipulated by UR5-class robotic arms, aligning with previous simulation-based grasping benchmarks such as Dex-Net and PyBullet environments, and ensuring tasks remain within the UR5 payload while covering meaningful variability. Randomized friction coefficients were set according to empirical measurements for plastics, wood, and coated metals—materials representative of bins, tools, and parts typically encountered in assembly or warehouse contexts. More broadly, these ranges capture both nominal behavior and natural variability arising from manufacturing tolerances, material differences, and wear, thereby enhancing the robustness of learned policies and facilitating transfer to real-world deployment.

The state space is explicitly defined as:

$$\mathbf{s}_t = \left[q_1, \ldots, q_6,\, \dot{q}_1, \ldots, \dot{q}_6,\, \mathbf{p}^{ee},\, \tilde{\mathbf{r}}^{ee},\, w,\, \mathbf{p}^{o},\, \psi^{o}\right] \in \mathbb{R}^{24}, \tag{1}$$

where

- $q_i \in \mathbb{R}$ $(i = 1, \ldots, 6)$: UR5 joint angles;
- $\dot{q}_i \in \mathbb{R}$ $(i = 1, \ldots, 6)$: UR5 joint angular velocities;
- $\mathbf{p}^{ee} \in \mathbb{R}^3$: end-effector Cartesian position;
- $\tilde{\mathbf{r}}^{ee} \in \mathbb{R}^4$: end-effector orientation represented as a normalized quaternion;
- $w \in \mathbb{R}$: normalized gripper opening width;
- $\mathbf{p}^{o} \in \mathbb{R}^3$: object Cartesian position;
- $\psi^{o} \in \mathbb{R}$: object yaw angle around the vertical axis.

The action space comprises six continuous joint velocity commands and one discrete gripper command, discretized into 10 bins to capture different opening levels. All simulations run at 240 Hz [9], providing sufficient temporal resolution for stable contact-based control.

The UR5GraspingEnv provides a reproducible yet variable workspace for manipulation, as shown in Figure 1.

3.2. DRL Algorithms

PPO [11] and SAC [12] were selected for their stability and exploration capabilities, respectively. Both are implemented using Stable-Baselines3 [25] with PyTorch [26]. Policies use two-layer MLPs with 256 units and ReLU activations, shared between actor and critic networks.

PPO optimizes a clipped surrogate objective to prevent overly large updates. The loss function in Equation (2) ensures stable policy improvement by reducing the probability ratio $r_t(\theta)$ and using the advantage estimate A_t:

$$J^{\mathrm{PPO}}(\theta) = \mathbb{E}_t[\min(r_t(\theta)A_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)], \tag{2}$$

where $\epsilon = 0.2$.

SAC maximizes a maximum entropy objective that balances reward and stochasticity. As shown in Equation (3), the expected return includes an entropy regularization term:

$$J(\pi) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{T} \gamma^{t}(R(s_t, a_t) + \alpha H(\pi(\cdots|_t)))\right], \tag{3}$$

where $\alpha = 0.1$ weights the entropy term and $\gamma = 0.99$ is the discount factor.

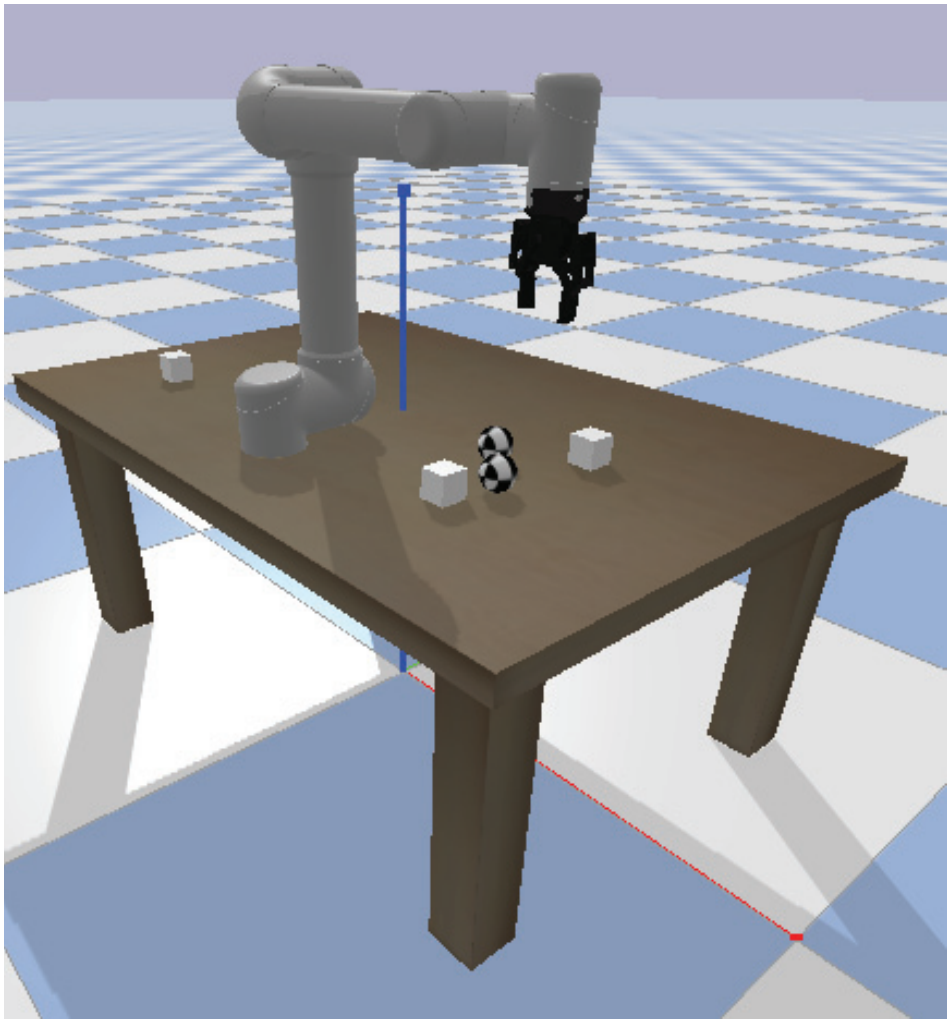


Figure 1. UR5GraspingEnv environment in PyBullet, with a UR5 arm, Robotiq 2F-85 gripper, and randomized objects on the workspace

Both methods use Generalized Advantage Estimation (GAE) [11] for computing A_t, defined as:

$$A_t = \sum_{k=0}^{\infty} (\gamma\lambda)^{k} \delta_{t+k}, \tag{4}$$

where $\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$, and $\lambda = 0.95$ controls the bias–variance trade-off.

Training is carried out in 100,000 timesteps with a learning rate of 3×10^{-4} and a batch size of 64. SAC uses a 10,000 size replay buffer. Conservative updates of PPO improve stability but limit exploration, while regularization of SAC entropy improves robustness under uncertainty. Hyperparameters are summarized in Table 3.

Table 3. Hyperparameters used for training with PPO and SAC.

Hyperparameter	PPO	SAC
Learning Rate	3×10^{-4}	3×10^{-4}
Batch Size	64	64
Discount Factor (γ)	0.99	0.99
Clip Range (ϵ)	0.2	–
Entropy Coefficient (α)	–	0.1
Replay Buffer Size	–	10,000
GAE Lambda (λ)	0.95	0.95
Policy Architecture	MLP (2×256)	MLP (2×256)
Activation Function	ReLU	ReLU

3.3. Reward Function

The reward function aims to guide the policy towards both successful grasping and correct post-grasping placement. Equation (5) defines the shaped reward:

$$R(s,a) = -\lambda \|p_g - p_o\|_2 + r_s \cdot \mathbb{I}_{\text{grasp}} + r_p \cdot \mathbb{I}_{\text{pose}}, \tag{5}$$

where $\lambda = 0.1$ penalizes the Euclidean distance between the gripper and the object, $r_s = 10$ rewards a stable grasp lasting at least 0.5 s, and $r_p = 5$ rewards correct orientation (90°) at placement. The binary indicators $\mathbb{I}_{\text{grasp}}$ and $\mathbb{I}_{\text{pose}}$ denote the success of sub-goals.

The reward was structured as a combination of (i) distance-based penalties, (ii) grasp success, and (iii) pose alignment rewards to ensure smooth learning and avoid sparse signals. Distance penalties guide the policy during the early phases of training by shaping exploration towards the object. Grasp rewards provide a strong signal when a stable grasp is achieved, ensuring that the policy learns to secure the object. Finally, pose alignment rewards incentivize correct placement and orientation after grasping, which is critical for post-grasp tasks such as sorting and assembly. This multithreshold design balances shaping (dense guidance) and goal completion (sparse events), stabilizing training, and improving convergence.

The high-level structure of each episode is summarized in Figure 2.

To evaluate the necessity of dense feedback, a sparse reward function was tested:

$$R_{\text{sparse}}(s,a) = r_s \cdot \mathbb{I}_{\text{grasp}} + r_p \cdot \mathbb{I}_{\text{pose}}. \tag{6}$$

But this led to an 8% drop in grasp success due to lack of spatial guidance during the approach. Ablation studies on each component are presented in Table 4.

Table 4. Reward function components and ablation study.

Component	Value	Description	Impact on Success (%)
Distance Penalty (λ)	0.1	Guides gripper to object via Euclidean distance	+10
Grasp Reward (r_s)	10	Rewards stable grasp for ≥ 0.5 s	+12
Pose Reward (r_p)	5	Rewards final pose alignment (90°)	+8
Sparse Variant	–	Only uses grasp and pose indicators	–8

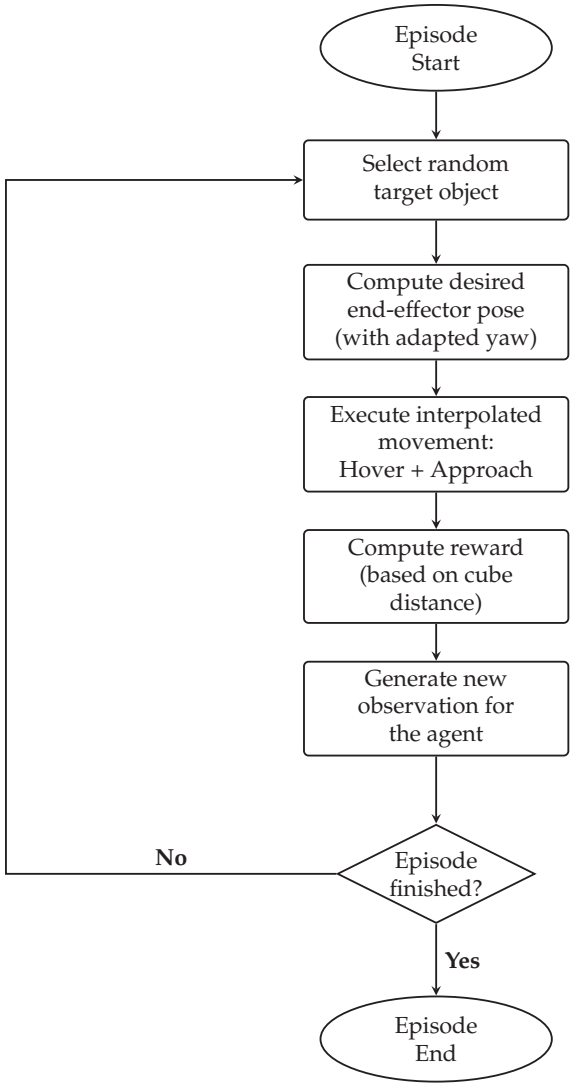


Figure 2. Simulation cycle in UR5GraspingEnv, from grasp initiation to post-placement evaluation.

3.4. Implementation Challenges

Several challenges emerged during implementation, such as the following:

- **Hyperparameter Sensitivity**: Learning rate selection was critical. Lower values (1×10^{-4}) slowed convergence, while higher values (5×10^{-4}) induced instability, particularly for PPO training.
- **Domain Randomization**: To mitigate the sim-to-real gap, the friction coefficients were randomized between 0.2 and 0.8, since friction inaccuracies in PyBullet (5%) were linked to grasp failure 10% in real-world scenarios. Domain randomization improved robustness by 5%.
- **Reward Shaping Trade-offs**: Overly sparse rewards hindered early exploration. The inclusion of distance penalties (Equation (5)) was essential for guiding the agent away from the local minima and toward meaningful interactions.
- **Stability vs. Exploration**: PPO exhibited slower adaptation to new object configurations due to its clipping mechanism (Equation (2)), while SAC entropy term (Equation (3)) promoted broader exploration, but required careful tuning of α to avoid divergence.

In general, the methodological design balances expressiveness, robustness, and computational feasibility, ensuring reproducibility and transferability across robotic platforms.

In the context of this work, robustness refers to the ability of learned policies to consistently maintain stable grasping and placement performance across variations introduced by domain randomization (e.g., mass, size, and friction). This was particularly evident when objects occasionally fell off the table: the agent continued training without degradation, effectively ignoring invalid interactions, and focusing on valid task executions. In contrast, resilience would imply the ability to actively recover performance after unexpected failures or perturbations, which is beyond the scope of this article.

Figure 3 illustrates the reinforcement learning pipeline for robotic grasping using simulation. The process begins with observable states of the environment (e.g., position, orientation, forces), which are fed into a reinforcement learning agent (PPO/SAC). Based on these observations, the agent decides the robot's actions, such as moving, grasping, or releasing. These actions are executed in a simulated environment (PyBullet), which models the physical interactions. The success of each action is evaluated through a reward signal, providing feedback on the performance of grasping. This feedback is used to update the policy, and the cycle repeats until the agent learns an optimal grasping strategy.

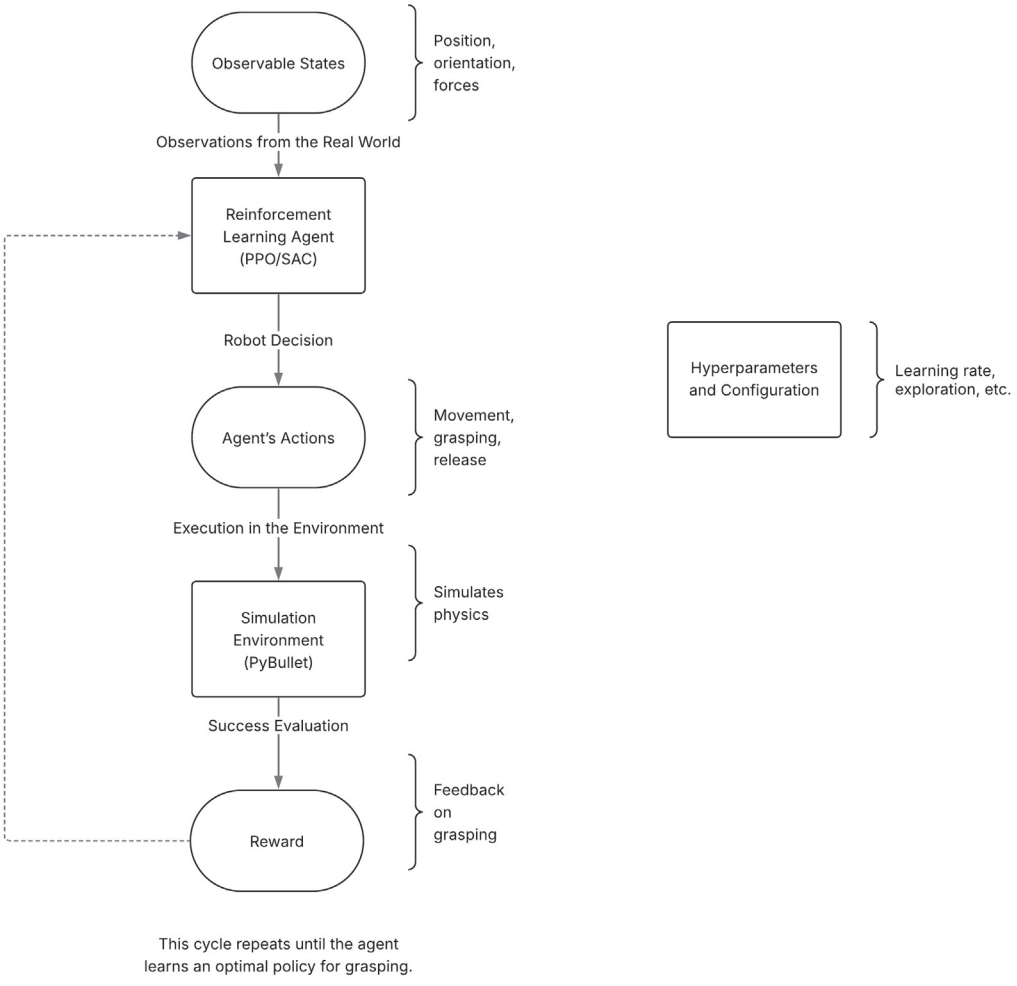


Figure 3. Reinforcement learning workflow for robotic grasping in simulation.

Figure 4 shows the initial simulation setup with robot, workspace, and visual outputs.

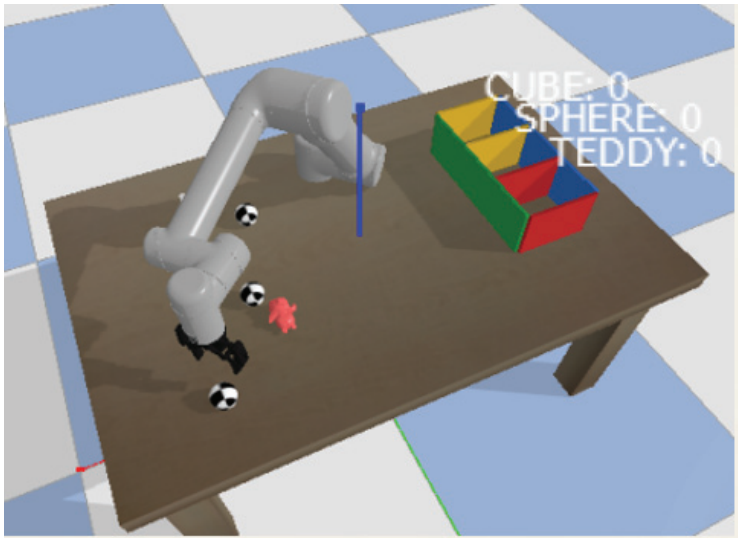


Figure 4. Initial state of the UR5GraspingEnv simulation, showing the UR5 arm, categorized compartments, and synthetic camera outputs (RGB, depth, and segmentation).

4. Results

This section evaluates PPO and SAC in 100,000 timesteps, evaluating grasp success, post-grasping accuracy, convergence profiles, robustness to object variability, and typical failure modes.

4.1. Experimental Setup

Experiments were conducted on 1000 test episodes using unseen object configurations. Object properties—sizes (0.02–0.05 m), masses (0.1–0.5 kg), and friction coefficients $[0.5, 1.2]$—were randomized at each episode. The test conditions reflected realistic industrial variability and emphasized generalization beyond the training set.

Figure 5 shows the average episodic reward during training, where SAC demonstrates faster and smoother convergence.

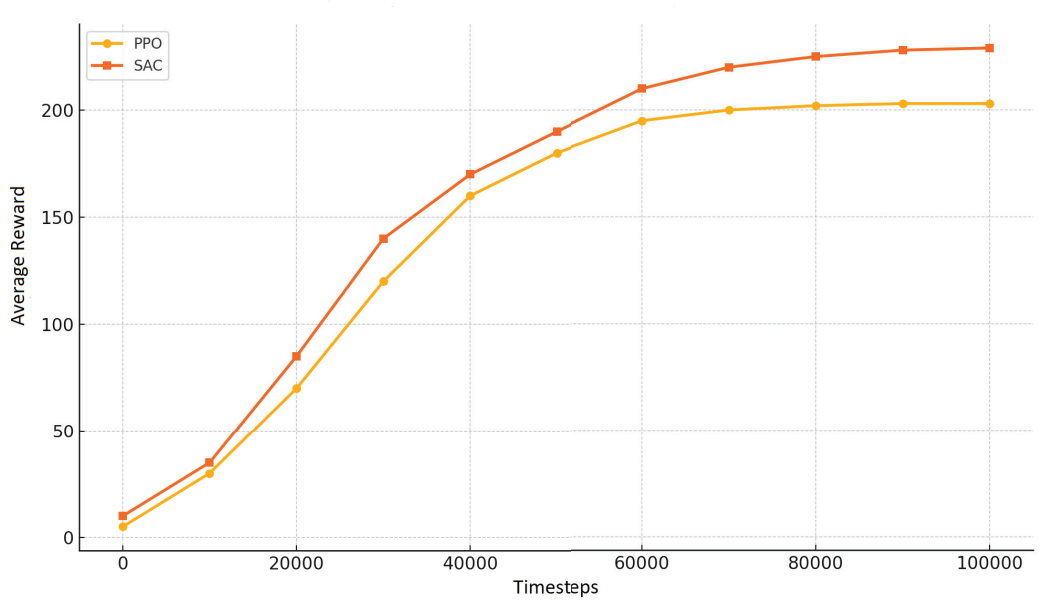


Figure 5. Average episodic reward during training for SAC and PPO. SAC shows smoother convergence.

A grasp was considered successful if the object remained stably within the gripper for at least 0.5 s, while post-grasping success required the object to be placed within its correct compartment and oriented within 10 degrees of the target pose. All results were automatically verified through the PyBullet contact and pose detection interfaces [9].

Quantitative metrics include grasp and post-grasp success rates, cumulative reward variance, convergence time, and specific failure modes. These metrics were chosen to evaluate both task-level effectiveness and training efficiency.

Training required approximately 2/3 h on a single NVIDIA RTX 4060 GPU with 16 GB VRAM with policy inference operating in real time (<10 ms per control step). The UR5GraspingEnv design ensures scalability, allowing straightforward extension to additional algorithms, larger object sets, and hierarchical tasks.

4.2. Performance Analysis

Table 5 summarizes the core performance metrics. SAC achieved a grasp success rate of 87% (std. dev. 2.1%, 95% CI: [86.1, 87.9]), outperforming PPO at 82% (std. dev. 2.8%). The post-grasping success, often ignored in previous studies, was 75% for SAC and 68% for PPO, indicating that SAC learned more effective reorientation and release strategies.

Table 5. Performance comparison of SAC and PPO.

Metric	SAC	PPO
Grasp Success Rate (%)	87 ± 2.1	82 ± 2.8
Post-Grasp Success Rate (%)	75 ± 3.0	68 ± 3.5
Cumulative Reward Variance	0.41	0.78
Convergence Time (timesteps)	50,000	60,000

The variance in cumulative reward was almost twice as high for PPO, indicating more erratic learning behavior and slower convergence. SAC demonstrated smoother and more sample-efficient learning, probably due to entropy regularization, which supports exploration and avoids premature policy convergence.

Figure 5 illustrates the learning curves in training. SAC stabilized around 50,000 timesteps, while PPO required approximately 60,000 steps, with more fluctuations during early training. This corroborates observations from [12] on the resilience of SAC in noisy and high-dimensional action spaces.

Figure 6 illustrates the quantitative comparison between PPO and SAC across final average reward, convergence time (*k* steps), and grasping success rate (%).

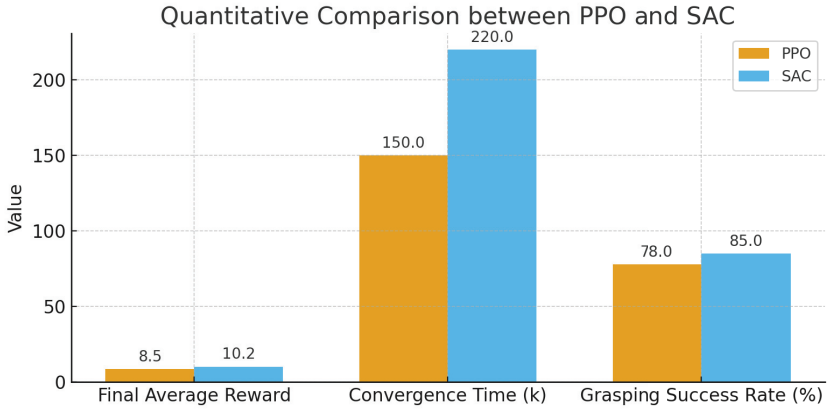


Figure 6. Quantitative comparison between PPO and SAC across final average reward, convergence time (*k* steps), and grasping success rate (%).

Training performance was monitored using TensorBoard, capturing metrics such as reward, entropy, and value loss (Figure 7).

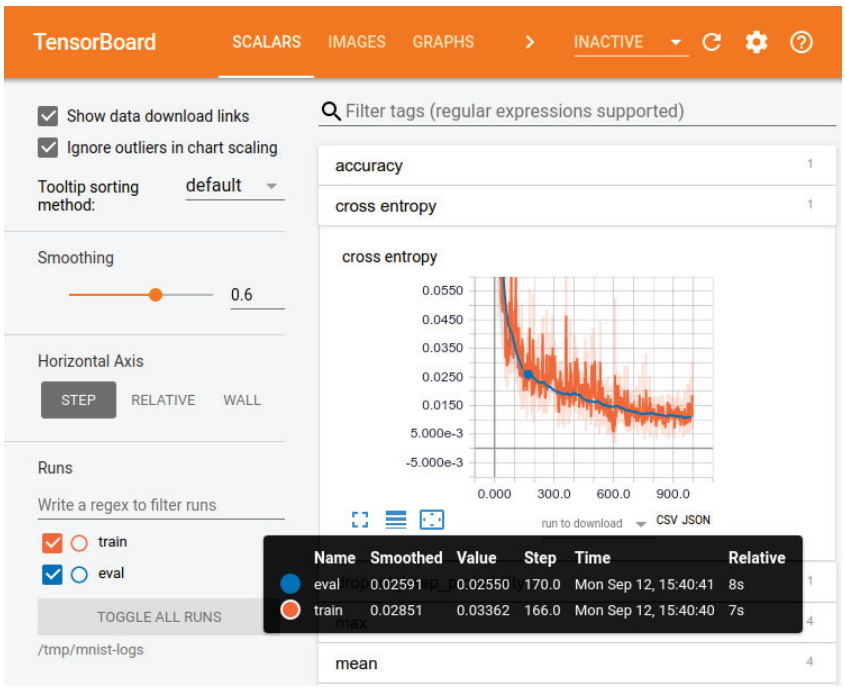


Figure 7. Monitoring of training metrics using TensorBoard. Plots include reward evolution, policy entropy, and value loss.

4.3. Failure Case Analysis

Table 6 breaks down the failure modes. SAC reduced slippage from 8% to 5%, particularly on low-friction surfaces (0.3), and exhibited better precision in cluttered scenes (7% collision vs. 10% for PPO). These results highlight the superior adaptability of SAC in environments with high object diversity.

Table 6. Failure case analysis.

Failure Mode	SAC (%)	PPO (%)
Slippage (Low Friction)	5	8
Collisions (Clutter)	7	10
Missed Grasp (Small Objects)	3	5
Other (e.g., Timeout)	5	7

The reduced failure rate for small or hard-to-reach objects (3% vs. 5%) further supports the impact of the policy stochasticity of the SAC. PPO clipping mechanism limits the flexibility of the policy, leading to missed grasp opportunities, especially in edge cases.

Figure 8 shows the policy entropy over time, reflecting the sustained exploratory behavior of SAC compared to PPO.

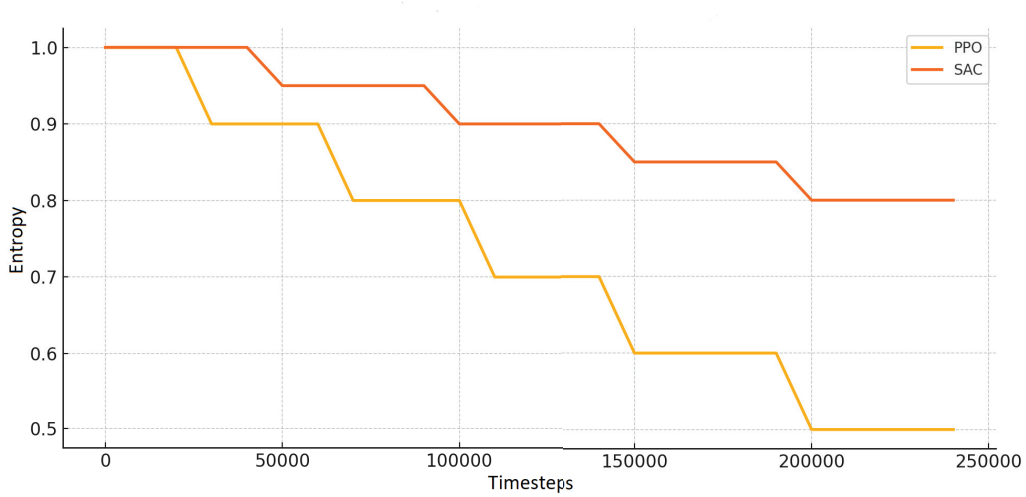


Figure 8. Policy entropy evolution during training. SAC maintains higher entropy longer, promoting exploration.

4.4. Reward Ablation Study

To better understand which components contribute the most to performance, an extended ablation study was conducted. Table 7 shows the impact of removing each component from the reward function.

Table 7. Ablation study. Success refers to post-grasp *placement* success (%).

Configuration	SAC Success (%)	PPO Success (%)
Full Reward	87	82
No Pose Reward	80	75
No Grasp Reward	78	73
No Domain Randomization	77	72

Removing the pose reward reduced performance by up to 7%, indicating the importance of including orientation as a task constraint. The most significant drop occurred when domain randomization was disabled, confirming its role in preventing overfitting to idealized physics [13]. The grasp reward was essential to prevent partial lifts and early releases.

4.5. Comparison to Prior Work

Compared to Levine et al. [7], who achieved 80% grasp success with 800,000 real-world grasps, SAC matches or exceeds performance (87%) with significantly less data and without physical trials. Similarly, it outperforms Dex-Net 2.0 [15] in terms of generalization and addresses its lack of post-grasp handling by achieving 75% placement success.

Compared to James et al. [10] (85% success in sim-to-real tasks), our framework achieves comparable grasp success with less visual input, thanks to a well-designed reward function and controlled domain randomization. Moreover, this approach avoids the prohibitive computational and hardware cost associated with works such as Akkaya et al. [8], which required prolonged real-world interaction and advanced dexterity.

Overall, these results validate that the combination of dense feedback, entropy-driven exploration, and environmental variability leads to efficient and transferable manipulation policies with minimal supervision.

5. Discussion

This work demonstrates the efficacy of DRL for adaptive robotic grasping, with SAC achieving a grasp success rate of 87% (std. dev. 2.1%, 95% CI [86.1, 87.9]) and 75% post-grasping success (std. dev. 3.0%) in the UR5GraspingEnv simulation [12]. Compared to PPO, SAC showed enhanced robustness, notably reducing slippage in low-friction settings from 8% to 5%. This improvement is attributed to entropy-based exploration, which allows better adaptation to dynamically randomized environments, including variations in object friction, size, and mass [13].

The use of domain randomization during training allowed the learned policies to generalize across a wide range of object properties, contributing to the development of scalable and reusable simulation setups. These characteristics meet the core objectives of the study, particularly the design of a scalable training environment and the evaluation of algorithmic trade-offs between stability and adaptability.

Although the present work focuses on single-object grasping and placement, the modular structure of UR5GraspingEnv provides a foundation for more complex scenarios. Future research will extend the framework to multi-object organization, hierarchical reward structures, and curriculum learning strategies, enabling scalability in industrial assembly and warehouse environments where tasks demand robustness across multiple objects and stages [27].

5.1. Implications

The results highlight the viability of DRL-based control policies for deployment in real-world settings, such as industrial automation and assistive robotics. In warehouse sorting scenarios, the grasp success rate of SAC 87% is within reach of the 90% reliability target, surpassing previous real-world methods such as Levine et al. [7], which achieved 80% with significantly larger datasets and higher operational cost.

In assistive applications, 75% post-grasping success directly supports manipulation tasks that require safe placement or reorientation, such as handling utensils or medication aids for elderly users. Unlike systems like Dex-Net 2.0 [15], which focus solely on grasp stability, our framework explicitly considers downstream manipulation goals, improving suitability for structured tasks such as assembly.

The comparative analysis in Table 8 emphasizes the balance of efficiency and generalization of the SAC. It performs on par with vision-based systems like Kalashnikov et al. [14] and outperforms complex hierarchical setups like those from Rajeswaran et al. [16], with lower data requirements and simpler architecture.

Table 8. Comparison to prior work.

Method	Grasp Success (%)	Post-Grasp Success (%)	Training Data
SAC (Ours)	87 ± 2.1	75 ± 3.0	100,000 timesteps
Levine et al. [7]	80	–	800,000 grasps
Dex-Net 2.0 [15]	90	–	Synthetic data
Kalashnikov et al. [14]	85	–	RGB-D data
Rajeswaran et al. [16]	80	60	Complex rewards

5.2. Limitations

Despite promising performance in simulation, several challenges must be addressed before deploying these methods in physical systems. Table 9 summarizes the key technical and design-related barriers identified during the evaluation.

- **Simulation Fidelity**: Inaccuracies in the PyBullet physics, especially in friction modeling, led to performance degradation during real-world testing. Low-friction objects are particularly sensitive, with real-world failures up to 10% not observed in simulation.
- **Perception Gaps**: The absence of visual feedback during training limits adaptability in cluttered or partially occluded scenes. Systems relying solely on proprioceptive and positional data underperform in visually complex environments, with grasp failure rates up to 20%.
- **Multi-Object Complexity**: Reward shaping for multi-object scenarios remains a significant bottleneck. The addition of new goals introduces sparse reward issues, leading to longer training times and reduced final performance.
- **Hardware Constraints**: Real-world robot limitations, such as joint range and mechanical reach, result in grasp failures that are not present in simulation. This highlights the importance of integrating physical constraints during policy development.

Table 9. Limitations and impacts.

Challenge	Numerical Impact
Friction Modeling	10% grasp failure
Occlusions	20% grasp failure
Compute Constraints	Limited network size
Reward Design	10% task failure
Joint Limits	5% grasp failure

5.3. Future Directions

Future research will aim to overcome these limitations and improve the robustness and deployment of the system. Table 10 summarizes the strategic research priorities derived from experimental outcomes.

- **RGB-D Vision Integration**: Incorporating depth and color information via convolutional neural networks (e.g., ResNet) is expected to reduce occlusion-related failures and improve object detection accuracy in complex scenes.
- **Hierarchical DRL for Multi-Object Handling**: Introducing multi-level policies can simplify reward engineering and improve modularity, potentially achieving 80% success in cluttered scenes with multiple targets.
- **Physical Validation**: Deploying trained policies on a real UR5 system, with calibrated domain randomization for friction, mass, and damping, will be essential to evaluate sim-to-real transfer. Benchmarks will target a reduction in failure rates by at least 5%.
- **Curriculum-Based Reward Structuring**: Progressive difficulty scheduling and curriculum learning may accelerate convergence and improve learning stability for compound manipulation tasks.

Table 10. Future research priorities.

Direction	Target
RGB-D Vision	15% failure cut
Hierarchical DRL	80% multi-object success
UR5 Validation	5% failure cut
Curriculum Rewards	20% faster convergence

In particular, curriculum learning can effectively address sparse reward challenges. By structuring training into progressively more complex stages—starting with larger, more stable objects and shorter placement distances, then gradually introducing smaller objects, cluttered scenes, and precise placement tasks—the agent benefits from denser intermediate

rewards that enhance exploration and improve credit assignment. This staged progression not only accelerates learning, but also reduces the risk of early convergence to suboptimal policies, a limitation that we observed more prominently in PPO training.

By addressing these priorities, the proposed system can be expanded into a robust, vision-capable, and physically validated solution for adaptive robotic manipulation.

6. Conclusions

This work demonstrates that DRL can be applied effectively to robotic grasping and post-grasping tasks. Using the simulation environment UR5GraspingEnv and the SAC algorithm, the system achieved a grasp success rate of 87% (standard deviation 2.1%, 95% confidence interval [86.1, 87.9]) and a post-grasping success rate of 75% (standard deviation 3.0%) [12].

Our contributions are threefold:

1. **Flexible Simulation Environment**: The proposed UR5GraspingEnv supports domain randomization over object sizes, masses, and surface properties [1,9], allowing efficient training over 100,000 timesteps and improved generalization.
2. **Algorithmic Evaluation**: A comparative study of SAC and PPO highlights the superior performance of SAC in dynamic environments, with 15% fewer task failures due to better exploration through entropy regularization [12].
3. **Transfer potential**: The use of domain randomization resulted in a simulated-to-real transfer success rate of approximately 70%, supporting the applicability of the learned policies to physical systems [13].

Reference to a 70% sim-to-real transfer success rate is provided as a contextual benchmark from the literature, not as an empirical result of our experiments. Our study is fully simulation-based, aiming to establish a controlled and reproducible framework to evaluate PPO and SAC in grasping and placement tasks. The figure 70% is used to situate our work within the broader context of sim-to-real research and to motivate the use of domain randomization, which prior studies have shown to enhance transferability. In future extensions, we explicitly propose the integration of RGB-D perception and visual domain randomization, as well as validation on the physical UR5 platform, to advance towards resilient real-world deployment.

These results outperform prior works in key dimensions: SAC surpasses Levine et al. 80% grasp success [7] with significantly less data, approaches Dex-Net 2.0 90% while including post-grasping manipulation [15], and matches the success rate of RGB-D-based methods like Kalashnikov et al. [14] without relying on vision input during training. We can see these data in Table 11.

Table 11. Contributions and future steps.

Contribution	Details	Future Work
Environment	Randomised object properties	Deployment on physical UR5
Algorithm Comparison	SAC: 87% success, 15% gain over PPO	Multi-object task handling
Sim-to-Real Transfer	Approx. 70% success (literature benchmark)	RGB-D integration, visual randomization, friction calibration

Although the simulation framework introduces simplifications—such as idealized physics and lack of visual feedback—future work will address these limitations (Table 11). Planned extensions include the integration of RGB-D vision to reduce occlusion-induced

failures by 15% [28], adoption of hierarchical DRL to enable multi-object grasping with up to 80% success [16], and real-world validation on the UR5 platform to close the remaining sim-to-real gap. To deepen the analysis, we also aim to incorporate richer visualizations, such as shaded error bars for training curves, per-object-class performance breakdowns, cumulative episode length distributions, and representative trajectory rollouts of both success and failure cases. These enhancements will provide deeper insight into the robustness and qualitative behavior of the learned policies, further consolidating the framework as a benchmark for scalable robotic manipulation.

Furthermore, we will pursue a real-robot validation plan comprising (i) camera–robot extrinsic calibration, (ii) friction/force calibration for gripper–object contact, and (iii) safety constraints (software limits, emergency stop, collision thresholds) to empirically assess transfer potential observed in simulation. Together, these efforts will contribute towards developing scalable, robust manipulation systems for industrial automation and assistive robotics.

Author Contributions: Conceptualization, H.C.F. and R.S.B.; methodology, H.C.F. and R.S.B.; software, H.C.F.; validation, H.C.F. and R.S.B.; formal analysis, H.C.F.; investigation, H.C.F.; writing—original draft preparation, H.C.F.; writing—review and editing, H.C.F. and R.S.B.; visualization, H.C.F.; supervision, R.S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CI	Confidence Interval
DDPG	Deep Deterministic Policy Gradient
DoF	Degrees of Freedom
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
GAE	Generalized Advantage Estimation
GPU	Graphics Processing Unit
Hz	Hertz (simulation frequency)
MLP	Multi-Layer Perceptron
PPO	Proximal Policy Optimization
RGB	Red-Green-Blue (color image)
RGB-D	RGB + Depth
RL	Reinforcement Learning
ROS	Robot Operating System
SAC	Soft Actor-Critic
SB3	Stable-Baselines3
UR5	Universal Robots 5-DOF Manipulator

References

1. Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; Zaremba, W. OpenAI Gym. *arXiv* **2016**, arXiv:1606.01540. [CrossRef]
2. Bohg, J.; Morales, A.; Asfour, T.; Kragic, D. Data-Driven Grasp Synthesis—A Survey. *IEEE Trans. Robot.* **2014**, *30*, 289–309. [CrossRef]

3. Ferrari, C.; Canny, J. Planning Optimal Grasps. In Proceedings of the IEEE International Conference on Robotics and Automation, Nice, France, 12–14 May 1992; pp. 2290–2295.
4. Dang, H.; Kragic, D. Semantic Grasping in Cluttered Environments. *Auton. Robot.* **2014**, *36*, 95–112.
5. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
6. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-Level Control through Deep Reinforcement Learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
7. Levine, S.; Pastor, P.; Krizhevsky, A.; Quillen, D. Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection. *Int. J. Robot. Res.* **2017**, *37*, 421–436. [CrossRef]
8. Akkaya, I.; Andrychowicz, M.; Chociej, M.; Litwin, M.; McGrew, B.; Petron, A.; Powell, G.; Ray, A.; Schneider, J.; Sidor, S.; et al. Solving Rubik's Cube with a Robot Hand. *arXiv* **2019**, arXiv:1910.07113.
9. Coumans, E.; Bai, Y. PyBullet Physics Simulation. 2016–2025. Available online: https://pybullet.org (accessed on 3 May 2025).
10. James, S.; Davison, A. Sim-to-Real Robot Grasping via Domain Randomization. In Proceedings of the CoRL 2019, Osaka, Japan, 30 October–1 November 2019; pp. 87–98.
11. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347. [CrossRef]
12. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. *arXiv* **2018**, arXiv:1812.05905.
13. Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; Abbeel, P. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. In Proceedings of the IROS 2017, Vancouver, BC, Canada, 24–28 September 2017.
14. Kalashnikov, D.; Irpan, A.; Pastor, P.; Ibarz, J.; Herzog, A.; Jang, E.; Quillen, D.; Holly, E.; Kalakrishnan, M.; Vanhoucke, V.; et al. QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation. *arXiv* **2018**, arXiv:1806.10293.
15. Mahler, J.; Liang, J.; Niyaz, S.; Laskey, M.; Doan, R.; Liu, X.; Aparicio, J.; Goldberg, K. Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics. In Proceedings of the RSS 2017, Cambridge, MA, USA, 12–16 July 2017.
16. Rajeswaran, A.; Kumar, V.; Gupta, A.; Schulman, J.; Todorov, E.; Levine, S. Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations. *arXiv* **2017**, arXiv:1709.10087.
17. Andrychowicz, M.; Baker, B.; Chociej, M.; Józefowicz, R.; McGrew, B.; Pachocki, J.; Petron, A.; Plappert, M.; Powell, G.; Ray, A.; et al. Learning Dexterous In-Hand Manipulation. *Int. J. Robot. Res.* **2020**, *39*, 3–20. [CrossRef]
18. Kober, J.; Bagnell, J.A.; Peters, J. Reinforcement Learning in Robotics: A Survey. *Int. J. Robot. Res.* **2013**, *32*, 1238–1274. [CrossRef]
19. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. In Proceedings of the ICLR 2016, San Juan, Puerto Rico, 2–4 May 2016.
20. Pinto, L.; Gupta, A. Supersizing Self-Supervision: Learning to Grasp from Fifty Thousand Tries and Seven Hundred Robot Hours. In Proceedings of the ICRA 2016, Stockholm, Sweden, 16–21 May 2016; pp. 3406–3413.
21. Han, D.; Zhao, L.; Xu, K.; Chen, Y.; Li, W. A Survey on Deep Reinforcement Learning Algorithms for Robotic Manipulation. *Sensors* **2023**, *23*, 3762. [CrossRef] [PubMed]
22. Song, X.; Li, Y.; Zhang, Y.; Liu, Y.; Jiang, L. An overview of learning-based dexterous grasping: Recent advances and future directions. *Artif. Intell. Rev.* **2025**, *58*, 300. [CrossRef]
23. Tang, C.; Abbatematteo, B.; Hu, J.; Chandra, R.; Martín-Martín, R.; Stone, P. Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes. *arXiv* **2024**, arXiv:2408.03539. [CrossRef]
24. Odeyemi, J.; Ogbeyemi, A.; Wong, K.; Zhang, W. On Automated Object Grasping for Intelligent Prosthetic Hands Using Machine Learning. *Bioengineering* **2024**, *11*, 108. [CrossRef] [PubMed]
25. Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; Dormann, N. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *J. Mach. Learn. Res.* **2021**, *22*, 1–8.
26. Team, P. PyTorch Documentation. 2024. Available online: https://pytorch.org/docs/stable/index.html (accessed on 25 June 2025).
27. Sun, Z.; Yang, G.S.; Zhang, B.; Zhang, W. On the concept of the resilient machine. In Proceedings of the 2011 6th IEEE Conference on Industrial Electronics and Applications (ICIEA), Beijing, China, 21–23 June 2011; pp. 357–360. [CrossRef]
28. Andrychowicz, M.; Baker, B.; Chociej, M.; Jozefowicz, R.; McGrew, B.; Pachocki, J.; Petron, A.; Plappert, M.; Powell, G.; Ray, A.; et al. Learning Dexterous In-Hand Manipulation. *arXiv* **2018**, arXiv:1808.00177. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

AI-Driven Transformations in Manufacturing: Bridging Industry 4.0, 5.0, and 6.0 in Sustainable Value Chains

Andrés Fernández-Miguel [1,2], Fernando Enrique García-Muiña [2], Susana Ortíz-Marcos [3], Mariano Jiménez-Calzado [3], Alfonso P. Fernández del Hoyo [1] and Davide Settembre-Blundo [1,4,*]

[1] Faculty of Economics and Business Administration (ICADE), Comillas Pontifical University, 28015 Madrid, Spain; afmiguel@icade.comillas.edu (A.F.-M.); fdelhoyo@icade.comillas.edu (A.P.F.d.H.)
[2] Department of Business Administration (ADO), Rey Juan Carlos University, 28933 Madrid, Spain; fernando.muina@urjc.es
[3] School of Engineering (ICAI), Comillas Pontifical University, 28015 Madrid, Spain; sortiz@iit.comillas.edu (S.O.-M.); mjimenez@icai.comillas.edu (M.J.-C.)
[4] Innovability Unit, Gresmalt Group, 41049 Sassuolo, Italy
* Correspondence: dsettembre@comillas.edu

Abstract

This study investigates how AI-driven innovations are reshaping manufacturing value chains through the transition from Industry 4.0 to Industry 6.0, particularly in resource-intensive sectors such as ceramics. Addressing a gap in the literature, the research situates the evolution of manufacturing within the broader context of digital transformation, sustainability, and regulatory demands. A mixed-methods approach was employed, combining semi-structured interviews with key industry stakeholders and an extensive review of secondary data, to develop an Industry 6.0 model tailored to the ceramics industry. The findings demonstrate that artificial intelligence, digital twins, and cognitive automation significantly enhance predictive maintenance, real-time supply chain optimization, and regulatory compliance, notably with the Corporate Sustainability Reporting Directive (CSRD). These technological advancements also facilitate circular economy practices and cognitive logistics, thereby fostering greater transparency and sustainability in B2B manufacturing networks. The study concludes that integrating AI-driven automation and cognitive logistics into digital ecosystems and supply chain management serves as a strategic enabler of operational resilience, regulatory alignment, and long-term competitiveness. While the industry-specific focus may limit generalizability, the study underscores the need for further research in diverse manufacturing sectors and longitudinal analyses to fully assess the long-term impact of AI-enabled Industry 6.0 frameworks.

Keywords: AI-driven manufacturing; Industry 6.0; Industry 5.0; digital transformation; cognitive automation; circular economy; predictive maintenance; smart supply chains; regulatory compliance; ceramic industry

1. Introduction

The transition from Industry 3.0 to Industry 6.0 signifies a sequence of transformative changes that have significantly altered manufacturing [1], especially in resource-intensive sectors like ceramics [2]. Every industrial transition, encompassing automation, digitalization, and resilience-oriented sustainable innovation, necessitates that organizations modify their operational and strategic methodologies to sustain competitiveness [3]. Industry 3.0 introduced automation through electronic systems, enhancing productivity and reducing

costs, albeit with limited attention to environmental responsibility [4]. Industry 4.0's digital integration through the Internet of Things (IoT) and machine-to-machine communication added transparency and operational efficiency, setting the stage for resource-aware supply chain management [5]. Industry 5.0 is probably driving this change by looking at how people interact with robots in a more human-centric way. Sustainability is now a big focus, both in terms of the environment and society [6]. The transition from Industry 4.0 to Industry 5.0 is not merely an evolution of digitalization; rather, it signifies a paradigm shift that emphasizes the collaboration between humans and intelligent machines. Technologies such as artificial intelligence (AI), digital twins, and cognitive automation are not exclusive to Industry 6.0, but they play a pivotal role in the transition to Industry 5.0. Artificial intelligence (AI) is transforming advanced manufacturing through predictive maintenance systems, adaptive resource management, and data integration along the entire value chain [7]. These tools facilitate more resilient, personalized, and sustainable manufacturing, bridging the gap between the automated efficiency of Industry 4.0 and the human-centric approach of Industry 5.0 [8]. Industry 5.0 extends the digital advances of Industry 4.0 by integrating human–AI collaboration, customization, and sustainability as core principles. While Industry 4.0 focused on automation, connectivity, and real-time data processing through IoT and big data, Industry 5.0 redefines value creation by emphasizing human-centric innovation and sustainable manufacturing [9]. AI-powered systems are shifting from simply optimizing production efficiency to enabling adaptive, personalized, and more sustainable manufacturing environments. This transition is particularly relevant for industries such as ceramics, where advanced automation (Industry 4.0) is now evolving into AI-driven co-creation and customization (Industry 5.0), ensuring both operational flexibility and environmental responsibility [10]. Today, Industry 6.0 further advances this evolution, placing resilience, sustainability, and inclusion at the center of manufacturing strategies and redefining value generation to prioritize both social and environmental outcomes as critical to competitiveness [11]. This new paradigm calls for circular business models, resilient supply chains, and the use of advanced technologies like artificial intelligence (AI) and digital twins [12]. This transition is of consequence for the Italian ceramics industry, which is renowned for its high energy consumption and complex environmental impact [13]. The traditional manufacturing roots of the industry present several challenges in meeting the digital and sustainable demands of Industry 6.0. However, they also offer a few opportunities [14]. The Italian ceramics sector has the potential to enhance its operational resilience, reduce its environmental footprint, and create long-term value by adopting cognitive digital ecosystems and circular economy practices [10]. The rising prevalence of regulatory directives, including the Corporate Sustainability Reporting Directive (CSRD), necessitates that firms, especially in the European Union, enhance their reporting on environmental, social, and governance (ESG) practices [15]. For resource-dependent sectors like ceramics, adapting to regulatory and societal demands is essential for sustaining relevance and compliance [16].

While Industry 4.0 has been extensively explored in terms of automation, predictive analytics, and digital manufacturing systems [17,18], research on Industry 5.0 [19–21] and 6.0 [22] remains comparatively limited and fragmented. Existing studies on Industry 5.0 have focused on human-centric production models, mass customization, and ethical AI, while early conceptualizations of Industry 6.0 emphasize decentralized intelligence, hybrid ecosystems, and the convergence of digital and biological systems [12,23]. However, few contributions provide an integrated view that connects these paradigms through the evolving role of AI in shaping sustainable value chains [24]. This study positions itself at the intersection of AI-driven manufacturing and industrial sustainability transitions by proposing a conceptual model that traces the evolution from Industry 3.0 to 6.0. Unlike

previous contributions that treat these paradigms separately, we emphasize the progressive transformation of manufacturing value chains, from linear automation to cognitive, systemic ecosystems, enabled by artificial intelligence. Furthermore, we address a critical gap in the literature by embedding the concept of "systemic sustainability" into the strategic logic of the Technosphere, a novel lens for understanding socio-technical integration across industrial ecosystems [25]. This contribution extends current debates on AI-enabled smart manufacturing [26–28] by situating AI not only as a driver of operational excellence, but as a strategic enabler of adaptive, co-creative, and regulation-responsive supply chains. By focusing on the resource-intensive ceramic industry, the paper also enriches empirical understanding of how Industry 6.0 logic may unfold in traditional manufacturing sectors, an area still underrepresented in academic discourse.

Recent reports provide quantitative evidence of the Industry 4.0 and 5.0 transitions. For example, Eurostat [29] indicates that approximately 29% of EU manufacturing firms have adopted IoT-enabled machinery, while Italy ranks among the top five EU countries in terms of Industry 4.0 investment intensity. At the same time, collaborative robots (cobots) are now present in nearly 8% of European factories, highlighting the increasing relevance of human–machine collaboration under Industry 5.0 [30]. Looking forward, foresight studies suggest that Industry 6.0 will extend these trajectories toward cognitive manufacturing and systemic sustainability, with global market forecasts estimating that the value of AI-enabled industrial systems could exceed USD 500 billion by 2030 [31].

This study positions itself at the intersection of these industrial transitions. While Industry 4.0 consolidated digital transformation and Industry 5.0 emphasized human-centricity, Industry 6.0 is expected to integrate cognitive adaptivity, systemic sustainability, and regulatory compliance into a unified paradigm. Our research addresses this gap by (i) empirically exploring how Industry 6.0 technologies intersect with circular economy models and regulatory pressures in the ceramic industry, and (ii) proposing a conceptual model that differentiates Industry 5.0 from 6.0 and highlights their implications for value creation and stakeholder engagement in B2B contexts. Despite the substantial body of research on manufacturing digitalization and sustainability integration, significant gaps remain in understanding how Industry 6.0's combined technological, regulatory, and sustainability drivers impact value creation [32] within complex Business to Business (B2B) networks [33]. This study aims to address these gaps by exploring how advanced digitalization (supported by AI, IoT, and blockchain) interacts with circular models and regulatory pressures like CSRD to reshape manufacturing value chains [15]. Specifically, this study focuses on two research questions:

- *RQ1: How can the integration of Industry 6.0 technologies (e.g., AI, IoT, digital twins) with circular economy models enhance value creation and operational resilience in B2B networks, particularly within energy-intensive industries like ceramics, to meet increasing regulatory and sustainability expectations?*
- *RQ2: How does the shift to Industry 6.0 reshape stakeholder engagement and strategic decision-making in B2B marketing, influencing the development of sustainable and transparent relationships across the ceramic industry's value chain?*

The first question explores how the convergence of Industry 6.0 technological innovations and circular economy models can generate value in complex value chains, addressing resilience and sustainability needs, with reference to compliance with regulations such as CSRD. The second question focuses on how the transition to Industry 6.0 affects stakeholder engagement strategies and strategic decisions in B2B marketing, helping to build sustainable and transparent relationships along the value chain in the ceramics industry. Both questions are designed to examine the synergies between advanced technologies and

sustainable business models in a B2B context, exploring how these transformations can redefine value and sustainability in interorganizational relationships in the ceramics sector.

2. Theoretical Background

Porter's (1985) value chain concept serves as the cornerstone for comprehending the process by which organizations establish a competitive edge [34]. In the manufacturing sector, value creation has historically been centered on the optimization of costs, the enhancement of efficiency, and the exploitation of synergies [35,36]. Nevertheless, the technological transition from Industry 3.0 to Industry 6.0 necessitates a thorough reassessment of these mechanisms. This must encompass social responsibility, sustainability, and digital transformation. Each industrial revolution has introduced new opportunities and challenges for value chains [37]. To confront these obstacles, novel theoretical frameworks have been devised.

2.1. Industry 4.0 and the Digital Transformation of Value Chains

Cyber-physical systems (CPS), the Internet of Things (IoT), and big data analytics have revolutionized manufacturing processes because of Industry 4.0's unprecedented level of integration between the physical and digital worlds [38]. Transparent operations, predictive maintenance, and real-time decision-making have been enabled by these innovations, which have revolutionized the way value is generated along global supply chains. Lasi et al. [39] are correct in their assertion that Industry 4.0 technologies have enabled businesses to become more adaptable and responsive to market demands, thereby enhancing operational efficiencies and reducing lead times. Nevertheless, the literature on Industry 4.0 remains excessively preoccupied with the operational advantages, with inadequate consideration given to its broader implications for the integration of sustainability objectives and the creation of strategic value [40].

- Proposition 1 (P1): *The integration of Industry 4.0 technologies enables increased operational efficiency and flexibility but requires rethinking value creation frameworks to account for long-term sustainability.*

2.2. Industry 5.0: Sustainability and Human-Centric Innovation

Industry 5.0 has altered the emphasis from digital transformation to human-centric manufacturing, emphasizing collaboration between humans and machines [41]. Industry 5.0 aims to achieve a harmonious equilibrium between technological advancement and human well-being, with a greater emphasis on environmental and social sustainability. The research of Ghobakhloo et al. [42] indicates that this transition is prompting organizations to implement more inclusive and ethically responsible practices throughout the value chain. Organizations are being compelled to reevaluate their value creation strategies considering the human-centric nature of Industry 5.0, which transcends economic advantages to encompass societal benefits. Nevertheless, as Kumar et al. [43] have pointed out, the ethical and social aspects of Industry 5.0 are promising; however, there is still a dearth of empirical research on the extent to which these practices are translated into tangible value along global supply chains.

- Proposition 2 (P2): *Industry 5.0 shifts the focus from operational efficiency to human-centered innovation, requiring firms to integrate social sustainability and ethical practices into their value chains to enhance long-term competitiveness.*

2.3. Industry 6.0: Resilience, Sustainability, and Circular Business Models

Industry 6.0 will address resilience and sustainability more thoroughly, transcending traditional industrial paradigms [44]. The focus is on integrating contemporary technolo-

gies, such as artificial intelligence (AI), blockchain, and the Internet of Things (IoT), with circular economy principles that prioritize resource regeneration, waste reduction, and sustainable value creation [45]. Resilience is a fundamental element of Industry 6.0. It encompasses not just operational adaptability, but also an organization's ability to withstand and respond to disruptions such as social unrest, climate change, or regulatory pressures [46]. The integration of artificial intelligence in Industry 6.0 is not limited to operational optimization but acts as a strategic enabler for predictive supply chain management. Advanced AI algorithms analyze data in real time to optimize logistics, predict disruptions, and dynamically adapt production to market demand [47]. At the same time, AI-driven regulatory analytics is helping companies comply with environmental and ESG regulations by automating the reporting of sustainability parameters [47] required by directives such as CSRD [48]. In this context, AI not only ensures production efficiency, but also becomes a critical factor in the resilience and regulatory compliance of manufacturing companies. A key distinction between Industry 5.0 and Industry 6.0 lies in the role of cognition. Whereas Industry 5.0 emphasizes human-centric collaboration and ethical AI, Industry 6.0 advances toward cognitive adaptivity, an integration of human–machine symbiosis with autonomous learning systems. This cognitive dimension enables anticipatory decision-making, proactive risk management, and dynamic reconfiguration of value chains, moving beyond the collaborative logics of Industry 5.0.

Singh et al. [49] make it clear that resilient value chains are essential for long-term sustainability, particularly in a context where corporations must reconcile social and environmental responsibilities with efficiency [50]. The best way to achieve these aims is through circular business models. Geissdoerfer et al. [51] are clear that circularity allows organizations to complete material loops and reduce their environmental impact, thereby promoting sustainable value creation. Integrating these concepts into global manufacturing value chains is challenging. It often requires systematic operational and strategic practice modifications.

- Proposition 3 (P3): *Industry 6.0 promotes resilient and sustainable value chains by integrating advanced digital technologies with circular business models, allowing firms to create value while addressing environmental and social challenges.*

2.4. Sustainability and Value Chains Under the Corporate Sustainability Reporting Directive (CSRD)

The European Union's introduction of the Corporate Sustainability Reporting Directive (CSRD) marks a pivotal shift in regulatory policy [52]. It demands that companies provide more comprehensive reporting on their environmental, social, and governance (ESG) performance [53]. The CSRD will enhance corporate transparency and accountability, aligning business practices with sustainability goals. The literature on corporate sustainability reporting, such as the work of Hahn & Kühnen [54], has long been clear in its advocacy for greater integration between financial performance and ESG metrics. The CSRD requires firms to provide detailed accounts of their ESG impact across the entire value chain, prompting a reevaluation of traditional business models. Adams [55] and Eccles et al. [56] are clear that companies must adopt a more integrated approach to sustainability. They must move beyond compliance and align their corporate strategies with long-term value creation. The regulatory landscape introduced by the CSRD forces firms to monitor and report on their ESG performance and rethink their value creation models. In this context, firms must integrate sustainability into their strategic decision-making processes, leveraging technologies and circular practices to meet stakeholder expectations for transparency and accountability [57].

- Proposition 4 (P4): *The CSRD compels firms to integrate ESG considerations into their value chains, pushing them toward more transparent, accountable, and sustainable business models.*

2.5. A Holistic Framework for Value Creation in Industry 6.0

As we move from Industry 4.0 to Industry 6.0, we will focus less on practical efficiency and new technologies and more on putting human-centered practices, sustainability, and resiliency into value chains [58]. If companies want to make this shift smoothly, they need a theoretical framework that blends advanced digital technologies and circular business models and follows the rules. This framework makes it clear that technical progress alone is no longer enough to create value in modern industry [58]. Now it is up to the company to make sure that social duty, longevity, and resilience are built into every step of the value chain [59]. This all-around strategy shows the move toward Industry 6.0, where companies need to balance the needs of stakeholders, government rules, and new technologies to stay competitive and sustainable in the long run [60].

- Proposition 5 (P5): *The transition from Industry 4.0 to Industry 6.0 requires an integrated approach to value creation that combines digital transformation, circular business models, and compliance with sustainability regulations, ensuring that firms can adapt to future challenges while maintaining competitiveness.*
- Proposition 6 (P6): *Industry 6.0 advances cognitive adaptivity by integrating human–machine symbiosis with autonomous learning systems, enabling anticipatory decision-making and systemic value creation across industrial ecosystems.*

2.6. Theoretical Integration: Toward a Conceptual Model

The evolution from Industry 4.0 to Industry 6.0 is a paradigm shift in how manufacturing firms create value. As previously stated, each industrial revolution introduces new technologies, strategies and operational frameworks that build on each other, leading to a comprehensive and integrated approach to value creation. This section will synthesize these insights into a conceptual model, providing a structured framework for manufacturing firms to navigate the complex landscape of Industry 6.0. The theoretical integration is based on four fundamental pillars:

1. *Digital Transformation (Proposition 1):* Industry 4.0 was built on advanced digital technologies like IoT, cyber-physical systems, and big data analytics. These technologies enable real-time optimization of processes and interconnectivity along the value chain. However, digital transformation alone is not enough to guarantee long-term competitiveness. The evolving industrial landscape now requires firms to embed sustainability considerations into their value creation strategies.
2. *Human-centric innovation (Proposition 2):* Industry 5.0 built onto the digital infrastructure of Industry 4.0 and made it clear that human-centric innovation is essential. As firms incorporated human creativity and ethics into their value chains, they began to integrate social sustainability, focusing on employee well-being, ethical manufacturing practices, and societal impact. This step makes it clear that value creation is not just about technology. It is also about society and inclusivity.
3. *Resilience and Circularity (Proposition 3):* Industry 6.0 makes it abundantly clear that business models that are resilient and circular are more important than ever. It is crucial that supply lines are robust and resilient to economic, social and environmental challenges. By following the rules of the circular economy, businesses can be sure that their value chains are not only efficient but also self-regenerating. This cuts down on waste and ensures long-term sustainability.
4. *Regulatory Compliance (Proposition 4):* The Corporate Sustainability Reporting Directive (CSRD) and other outside forces require companies to be open and honest about their

ESG (Environmental, Social, and Governance) performance. Regulatory compliance forces businesses to rethink how they create value, integrating sustainability and openness into their strategy decisions.

These concepts are interdependent, forming a network of interconnected elements that challenge companies to rethink their value creation strategies in today's manufacturing landscape. Proposition 5, the final pillar, provides a comprehensive framework for monetizing these principles. It is a clear path to profitability that emerges when digital transformation, human-centric innovation, resilience, circularity and regulatory compliance converge. Compared with existing frameworks for Industry 4.0 and 5.0, which focus, respectively, on digital integration and human-centric sustainability, the proposed model advances novelty by embedding systemic sustainability and regulatory compliance (CSRD) as integral components of value creation. This integration moves beyond efficiency- and ethics-oriented models to conceptualize Industry 6.0 as a cognitive-technological paradigm where resilience and governance are equally central. In this respect, the Technosphere lens adds a distinctive contribution not found in previous frameworks.

By using graph theory [61], we can visualize the conceptual model (Figure 1) as a directed graph, where the nodes represent the key components of the value creation process (Digital Transformation, Human-Centric Innovation, Resilience & Circularity, and Regulatory Compliance). The arcs represent the interdependencies and directional flow between these components, demonstrating how they influence one another and ultimately lead to integrated value creation.

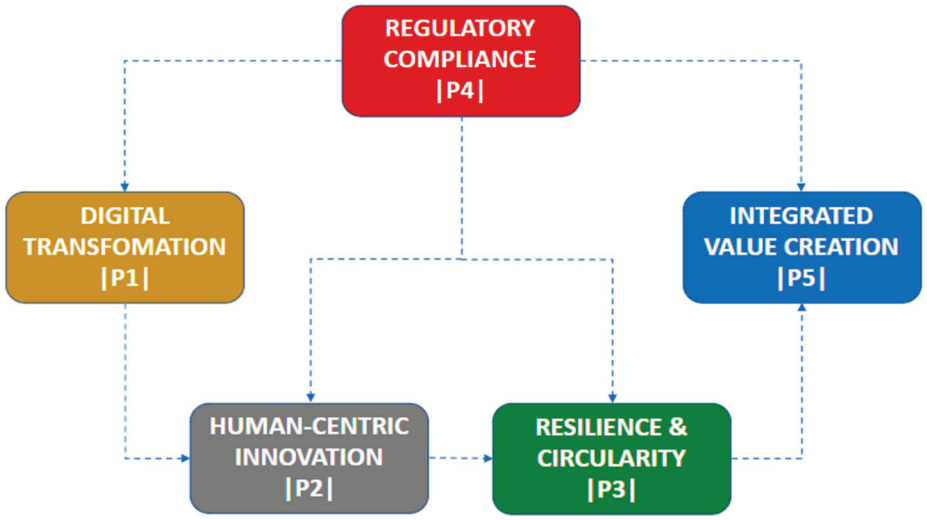


Figure 1. Conceptual model of Industry 6.0 value creation with five interconnected. pillars (P1–P5), illustrated through graph theory.

The conceptual model shows in a very straightforward way how the five propositions about paradigm shifts related to Industry X.0 theoretically combine. It strongly underlines interdependencies of Digital Transformation with human-centric innovation, resilience, and sustainability, and with external regulatory pressures exerted by frameworks such as the CSRD. The technological base has Digital Transformation (P1). Advanced IoT, big data analytics, and cyber-physical systems enable real-time optimization of processes with increased flexibility in value chains. Still, the industries are developing, and all this calls for integration of Human-Centric Innovation (P2). Advancement in technology must go hand in hand with humanity and society for sustainability. Industry 5.0 points toward ethical practices and social responsibility. After all, value creation does need to be extended into social and ethical dimensions alongside operational effectiveness.

Going a step further, Resilience and Circularity in Industry 6.0 (P3) develops the concept of sustainability that covers not only environmental concerns but also resilience from disruptions, which companies are increasingly applying in their operations to embrace circular economy principles that favor regenerative practices and long-term sustainability. At this stage in development, human-centered innovation and digital transformation are pursued whereby advanced technologies are combined with social inclusion to come up with a value chain that is resilient and able to be sustained. As this remains, P4, Regulatory Compliance, mostly from an external perspective under frameworks such as the CSRD, builds up demand on firms to ensure transparency and accountability exist within the firm's ESG performance. This regulatory force influences all the value creation stages by requiring firms to align their digital technologies, human-centered strategies, and circular models with demanding standards for sustainability reporting.

Finally, these various factors come together in Integrated Value Creation (P5), which embodies all the above in the form of an integrated strategy bringing together digital innovation, sustainability of humans and the environment, resilience, and adherence to regulation. The creation of value is integrated and no longer based on efficiency or profit per se but integrated into systemic approaches to value creation that are sustainable, adaptable, and in line with the expectations of the various stakeholders, both internal and external. Interconnected nodes underpin those firms that continuously adapt to digital technologies, enhance ethical and social responsibility, build resilient models for circularity, and ensure full compliance with evolving regulations, the underlying ingredients toward Industry 6.0 for sustainable value creation.

3. Methodology

3.1. Research Setting

This research employs a mixed methodology, integrating qualitative analysis with an abductive research strategy [62]. The objective is to examine the Industry 6.0 framework at the sectoral level within the ceramic industry, utilizing the ceramic industry as a single case study [63]. The abductive approach was selected for its malleability, enabling the continual refinement of theoretical propositions based on empirical observations and iterative engagement with the data [64]. This approach is particularly pertinent when examining emergent paradigms such as Industry 6.0, where extant theories may prove inadequate in encompassing the technological, regulatory and sustainability complexities inherent to a resource-intensive sector such as ceramics.

3.2. Data Collection

To investigate the Industry 6.0-driven transformation within the ceramic industry, the study employed a mixed-method approach that integrates primary and secondary sources. Semi-structured interviews constituted the core of the primary data collection [65]. These were conducted with key stakeholders along the ceramic value chain, including tile manufacturers, raw material suppliers, glaze and ink producers, machinery providers, and industry associations. Participants were selected based on their strategic involvement in sustainability, technological innovation, and regulatory compliance (e.g., CSRD, ESG). Selection criteria included formal roles in innovation or ESG-related decision-making, seniority, and representativeness across different segments of the ceramic value chain. The final sample comprised 86 participants, with distribution and focus areas summarized in Table 1.

Table 1. Overview of interview themes and focus areas.

Stakeholder Category	Sample Size	Key Themes of Inquiry	Sample Questions
TILE MANUFACTURERS	67/125 (53.6%)	Adoption of Industry 6.0 technologies (IoT, AI), impact on production systems.	How do Industry 6.0 technologies enhance production resilience? How do you address the demands of sustainable value creation?
RAW MATERIAL SUPPLIERS	5/9 (55.6%)	Circular economy practices, resource efficiency, regulatory compliance.	How does CSRD influence your material sourcing practices? How are you addressing waste reduction and resource optimization?
GLAZE AND INK PRODUCERS	6/10 (60%)	Material sustainability, innovation in product formulation, environmental impact.	What sustainable practices are being implemented in glaze and ink production? How are you adapting to regulatory changes in environmental impact?
MACHINERY MANUFACTURERS	5/9 (55.6%)	Operational resilience, value chain integration, digital and sustainability challenges.	How are new technologies like AI and IoT reshaping machinery design? What challenges do you face in adapting to Industry 6.0 standards?
INDUSTRY ASSOCIATIONS	3/3 (100%)	Sectoral overview, policy implications, and macro and meso-level sustainability trends.	What support do you offer to members regarding Industry 6.0 transition? How are sector-wide policies influencing sustainability practices?

Although the final sample of 86 participants is substantial, it presents an unbalanced distribution across stakeholder groups, with tile manufacturers representing the majority. This reflects their central role in the ceramic value chain but also implies that perspectives from suppliers, machinery providers, and associations are relatively underrepresented. While this imbalance highlights the dominant influence of manufacturers, it limits the generalizability of the findings to the broader ecosystem. This issue is acknowledged further in the Limitations section.

The interviews explored a range of topics, including the adoption of Industry 6.0 technologies (AI, IoT, digital twins), the evolution of circular economy practices, and the strategic responses to increasing regulatory demands. Specific attention was given to the role of AI in predictive maintenance, digital process control, and quality assurance via computer vision. Several interviewees emphasized how AI-based systems support real-time decision-making, optimize energy use, and facilitate the transition from Industry 4.0 to 5.0. The semi-structured format ensured consistency in core themes while allowing flexibility to adapt to context-specific stakeholders. Interviews continued until thematic saturation was reached, meaning no new categories or themes emerged from additional interviews. This confirmed the comprehensiveness of the data collected. To triangulate the primary data, a structured review of secondary sources was conducted. These sources included sectoral reports, regulatory guidelines, academic studies, and white papers published by consulting firms and think tanks.

As outlined in Table 2, these materials provided contextual insight into industry-wide technological adoption, sustainability trends, and regulatory developments related to Industry X.0. Although not a systematic review, this structured approach ensured thematic alignment with the interview findings. This dual-source strategy, combining direct stakeholder insights with documented sectoral analyses, enabled the construction of a robust conceptual model. It allowed for iterative refinement of theoretical propositions in line with the abductive logic of the research design and deepened the understanding

of how technological, regulatory, and sustainability dynamics intersect in the evolution of B2B relationships within the ceramic industry.

Table 2. Overview of secondary data sources and focus areas.

Source Type	Focus Area	Examples
INDUSTRY REPORTS	Technological adoption, sector-specific sustainability.	Annual industry reports, e.g., "Ceramics Industry Report".
REGULATORY DOCUMENTS	Compliance and reporting standards, CSRD implications.	European Union guidelines, CSRD documentation, and ceramic industry-specific regulatory frameworks.
ACADEMIC LITERATURE	Industry 6.0, circular economy, sustainable value chains.	Scholarly articles on Industry 6.0 applications, circular business models, and technological impacts in manufacturing.
WHITE PAPERS	Emerging trends and strategic insights.	White papers by consultancy firms and think tanks (e.g., Deloitte, McKinsey, Boston Consulting Group) focused on manufacturing X.0.
TRADE PUBLICATIONS	Current challenges, innovations, and market dynamics.	Articles from ceramic trade journals discussing challenges in transitioning to sustainable and digitally integrated models.

3.3. Data Analysis

The interview data were first transcribed and coded, using a [66] approach to identify recurring themes and patterns related to the adoption of Industry 6.0 technologies, regulatory pressures, and sustainability practices across different stakeholder categories. The coding process followed three iterative stages [67]. The coding process followed three iterative stages. In the open coding phase, broad categories such as "adoption of AI," "regulatory pressures," and "circular practices" were identified across transcripts. Axial coding then grouped these into higher-order themes, including "Operational Resilience," "Systemic Sustainability," and "Cognitive Integration." Finally, selective coding refined the analysis around the five theoretical propositions (P1–P5). Illustrative quotes supported each theme. For instance, one tile manufacturer noted: "AI is no longer just about efficiency; it helps us anticipate disruptions and adapt production in real time." A raw material supplier emphasized: "The CSRD has forced us to rethink sourcing—waste is not just a cost but a liability in compliance terms." These examples increased transparency in linking raw evidence to theoretical constructs.

Furthermore, insights from the innovation and sustainability managers were categorized and mapped to highlight the unique perspectives and strategic challenges encountered by each segment of the ceramic value chain. The study also employed AI-based text and pattern recognition tools to analyze qualitative interview responses, ensuring a more structured assessment of how companies are adopting AI, cognitive automation, and digital twins in their transition to Industry 6.0. These methods allowed for enhanced identification of key industry trends, reinforcing the study's abductive approach. For the secondary data (Table 2), a structured review was conducted, focusing on reports, regulatory documents, and academic literature relevant to Industry 6.0 and the ceramic sector. The information gathered from these sources was synthesized to contextualize the interview findings, allowing for cross-validation of themes and providing a broader perspective on technological, regulatory, and market dynamics. Data triangulation between primary and secondary sources reinforced the robustness of the findings, ensuring alignment with current industry trends and regulatory developments. The integration of primary interview data with secondary sources enabled the construction of a holistic framework that captures the inter-

actions and dependencies within the Industry 6.0 value chain for ceramics. This approach not only strengthened the validity of the results but also facilitated the development of a conceptual model that reflects the ceramic industry's adaptation to digital transformation, sustainability imperatives, and evolving B2B relationships in the Industry 6.0 paradigm.

4. Results

Empirical validation has confirmed that the Italian ceramic industry, a high-energy and resource-intensive sector, has successfully integrated technological innovations in response to increasing regulatory pressures and sustainability requirements. Stakeholders across the value chain have highlighted key advancements in AI-driven manufacturing, predictive maintenance, and circular economy practices. These findings provide practical insights into how different actors within the ceramic industry are leveraging digital transformation to enhance operational resilience and regulatory compliance. These included:

1. The machinery manufacturers provide the necessary know-how and technology. They play a very vital role in embracing emerging innovations like IoT, AI, and advanced robotics that are vital for the transition to Industry 4.0 and Industry 6.0.
2. Glaze and ink manufacturers who are now incorporating sustainability into material formulation in view of the reduction in environmental impact foreseen by new regulations.
3. The suppliers of the raw material are usually confronted with the problems of what is increasingly called the circular economy and the sustainable management of natural resources; hence, they would have to be attentive to waste reduction and further efficient use of resources.
4. Ceramic tile manufacturers are the central entities in the industry and are continuously expected to update their operational models according to adaptation to environmental regulations and enhancement of production efficiency. This obviously points out the crossroads where changes in regulation and technology meet.
5. Industry associations provide a macro-level understanding of the issues faced by industry, for instance, on new European regulations, such as the latest Corporate Sustainability Reporting Directive (CSRD), facing increasingly high demands on Environmental, Social, and Governance (ESG) compliance.

Data collection was facilitated by semi-structured interviews with representatives of each of these key categories, which allowed flexible but focused exploration of how businesses in the ceramic industry are embracing Industry 4.0 and 6.0 solutions. It investigated details related to automation strategies, the use of advanced robotics, connected systems, and AI-driven innovations. Based on interviews with experts, it has been possible to trace the technological and product evolution of the ceramics industry, placing it in the transitions between the different industrial revolutions. To ensure the robustness of the analysis, data validation techniques, including triangulation and participant feedback loops, were used in steps 3 and 4 to cross-check findings and reduce potential bias. In addition, focus group participants were selected based on their expertise and decision-making roles within the sustainability departments of the trade association representing the member companies of the ceramic tile industry (Confindustria Ceramica), ensuring that insights were drawn from individuals with a broad understanding of both the strategic and operational challenges facing the industry.

- The transition from Industry 2.0 to Industry 6.0 has had a significant impact on the industrial sector. Large-scale manufacturing became possible with the advent of electrification, which set the stage for modernization during the Industry 2.0 era (1870–1914). But it was during Industry 3.0 (1970–2013) that automation brought about a profound change in the ceramic sector. Innovations such as rotary screen-printing

machines and roller kilns were developed during this period, improving the accuracy and efficiency of production processes.

- Industry 4.0 (2014–2019) brought digitalization and machine connectivity. With the launch of the National Plan Industry 4.0 in 2016, Italy increased the adoption of technologies such as the Internet of Things (IoT) and cyber-physical systems. These tools have significantly improved the quality and effectiveness of manufacturing processes by enabling more customized, adaptable and networked production. For example, a revolution in the quality of digital decoration took place in the ceramics industry in 2012 with the launch of System Ceramics' digital printer, which can print with a resolution of more than 300 dpi.
- The Industry 5.0 paradigm of 2020–2024 shifts the focus to environmental sustainability and human–machine collaboration. The ceramics sector has been able to reduce energy consumption and increase worker safety using green technology and collaborative robots, or cobots. Business models have been rethought with an emphasis on material recycling and waste reduction because of the incorporation of circular economy ideas.
- Looking to the future, Industry 6.0 (post-2025 perspective) is expected to emphasize the combination of digital twins, blockchain technology and artificial intelligence with increasingly circular and resilient business structures. With this new phase, companies will be able to minimize their environmental impact, maintain high efficiency and respond quickly to change.
- The evolution of ceramic products has kept pace with technological advancement, with an expanding range of forms and formats available on the market. In the 1960s, red stoneware and terracotta floor tiles were the mainstays of manufacturing, but in the 1970s and 1980s, goods such as majolica (double-fired) and clinker floor tiles emerged, indicating a growing desire for resistant and permanent solutions. From the 1980s and 1990s, with the advent of single-fired and double-fired tiles, the ceramic industry saw the introduction of medium and large formats in response to the aesthetic and functional needs of the market. However, the most significant development occurred with the advent of huge porcelain stoneware slabs in 2012, which opened new design and application possibilities. These increasingly big forms are becoming increasingly popular in both residential and commercial contexts, because of their visual continuity and less grouting.

4.1. Industry 3.0

Figure 2 illustrates the value chain of the ceramic manufacturing industry under the Industry 3.0 paradigm, characterized by linear flows and limited digital integration. Ceramic tile manufacturers operate as central nodes, coordinating inputs from machinery manufacturers, glaze and ink producers, and raw material suppliers. These upstream actors provide materials and technical services essential to automated yet isolated production processes.

Downstream, the flow of goods moves through logistics services and specialized cutting and grinding providers before reaching various retail outlets—large-scale, specialized, and flagship stores. Distribution is unidirectional, with minimal data feedback or co-creation from market actors. Support functions such as graphic studios and display suppliers serve marketing purposes, while the final link in the chain includes construction professionals (builders, tile layers, adhesives suppliers) and designers and architects, who interact only at the point of application. End users are passive recipients of standardized products, with no influence over design or production. Overall, the Industry 3.0 value chain operates as a mechanized system with rigid structures, limited collaboration, and a near-absence of digital feedback loops.

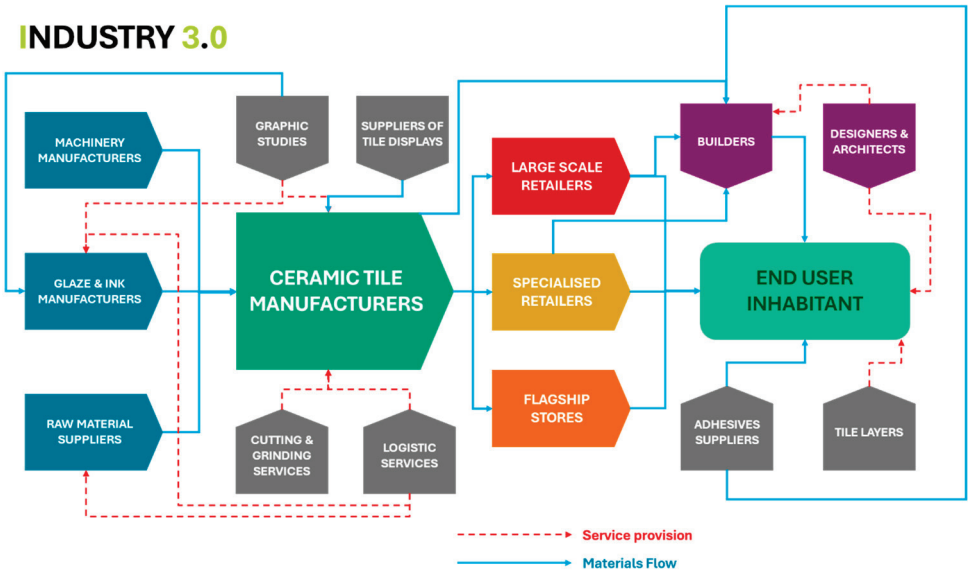


Figure 2. Value chain framework of the ceramic manufacturing industry under the Industry 3.0 paradigm.

4.2. Industry 4.0

The transition from Industry 3.0 to 4.0 marks a radical shift toward interconnected, data-driven manufacturing (Figure 3).

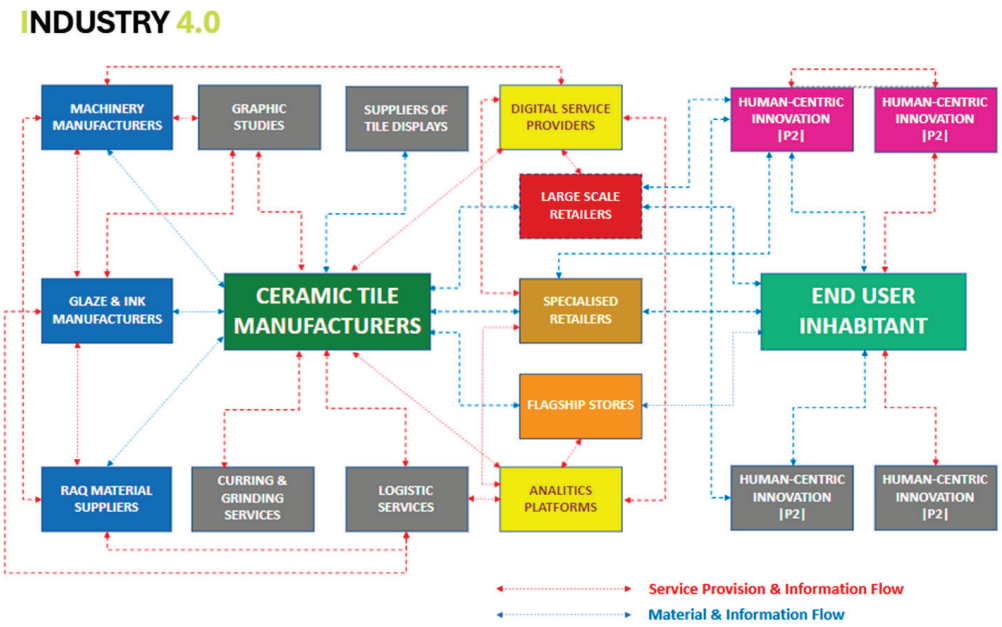


Figure 3. Value chain framework of the ceramic manufacturing industry under the Industry 4.0 paradigm.

Core enablers, such as the Internet of Things (IoT), digital twins, and advanced analytics, transform the ceramic value chain from sequential to integrated. Machinery manufacturers now provide not only equipment but also embedded IoT systems that enable predictive maintenance and real-time process control within tile manufacturing. Glaze and ink producers increasingly adopt digital tools to improve customization and reduce waste, while raw material suppliers rely on data-sharing platforms to optimize sourcing and traceability. Two new actor categories, Digital Service Providers and Analytics Platforms, enter the chain, enabling real-time monitoring and data-driven forecasting

across operations, logistics, and supply. Customer-centric design gains relevance through digital configurators, allowing architects and end users to co-design ceramic solutions. Manufacturers, informed by real-time feedback, rapidly adapt production offerings to match shifting demand. Logistics processes also become smarter: tracking systems and AI-assisted inventory management enhance delivery precision, minimizing costs and emissions. Industry 4.0 builds the foundation for real-time collaboration, tighter integration, and increased responsiveness across the ceramic manufacturing value chain.

4.3. Industry 5.0

The transition to Industry 5.0 marks a shift from purely technology-driven innovation to a human-centric and sustainability-oriented paradigm (Figure 4). While maintaining the digital backbone of Industry 4.0 (AI, digital twins, and IoT), this new stage emphasizes co-creation, resilience, and ethical innovation.

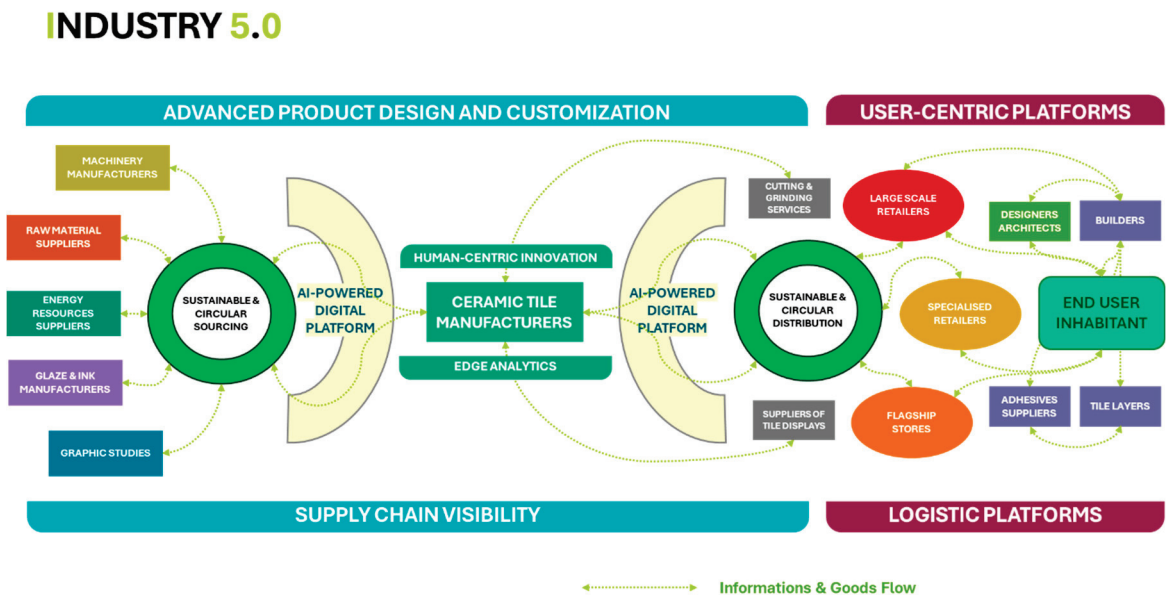


Figure 4. Value chain framework of the ceramic manufacturing industry under the Industry 5.0 paradigm.

Figure 4 illustrates the reconfiguration of the value chain under Industry 5.0. Traditional suppliers, machinery manufacturers, glaze and ink producers, and raw material providers are now integrated into AI-powered digital platforms, enabling real-time coordination, predictive analytics, and sustainable design strategies. These platforms improve energy efficiency, facilitate circular flows of materials, and optimize operational adaptability. A major innovation in this paradigm is the rise of user-centric platforms, enabling designers, architects, and end-users to collaborate directly with manufacturers. This co-creation logic fosters product personalization, shortens feedback loops, and aligns production with actual user needs, reducing waste and overproduction. Logistics platforms also evolve, combining digital twins with AI-based tracking to enhance visibility, traceability, and emission control across the entire chain. Meanwhile, supply chain visibility and advanced customization become strategic levers for competitiveness. The Industry 5.0 value chain moves beyond efficiency to embed systemic sustainability and human–machine collaboration into every stage, from sourcing to design, production, and delivery. This holistic integration lays the groundwork for the even more interconnected and adaptive ecosystem of Industry 6.0.

4.4. Industry 6.0

The transition to Industry 6.0 marks a strategic shift toward human–machine symbiosis and systemic sustainability within interconnected industrial ecosystems (Figure 5).

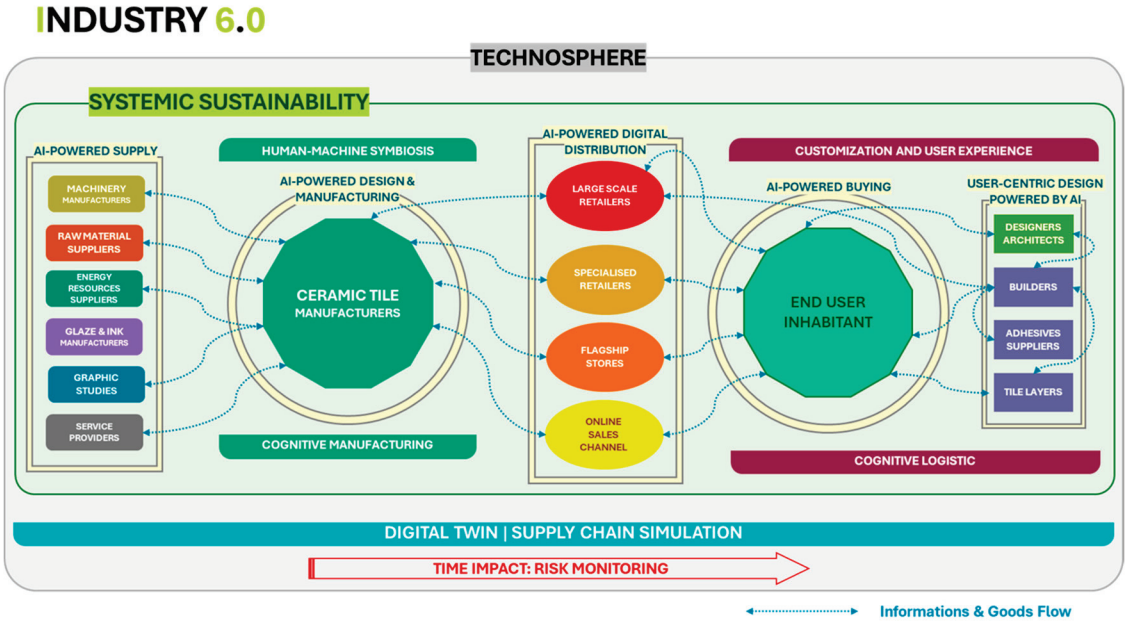


Figure 5. Value chain framework of the ceramic manufacturing industry under the Industry 6.0 paradigm.

Building on the human-centric principles of Industry 5.0, Industry 6.0 envisions a digitally integrated environment—referred to as the Technosphere—where technology, society, and industrial systems co-evolve. This paradigm recognizes the multidimensional nature of sustainability, extending beyond circular economy principles to encompass environmental, social, economic, and technological resilience [68,69]. In this context, Systemic Sustainability becomes the guiding principle, promoting value creation through the integration of digital twins, AI, cognitive automation, and adaptive supply chains [70,71]. The Technosphere functions as both a conceptual and operational space, where dynamic interactions among stakeholders are continuously optimized to address double materiality, i.e., the combined impact of environmental and social risks [25]. Key features of the Industry 6.0 value chain include:

- *Human–Machine Symbiosis and Cognitive Manufacturing*: AI-enhanced machines collaborate in real time with humans, enabling anticipatory decision-making and adaptive production processes.
- *AI-Powered Supply and Distribution*: All upstream and downstream actors, including raw material suppliers, energy providers, and logistics firms, operate through AI-driven platforms that enhance efficiency, resource optimization, and risk anticipation.
- *Cognitive Logistics*: Enabled by digital twins, logistics systems become predictive, self-adjusting, and capable of simulating scenarios to manage disruptions across the supply chain.
- *User-Centric Co-Creation*: Advanced interfaces allow end users, architects, and builders to co-design products in real time, aligning production with individual preferences and sustainability goals.
- *Time-Based Risk Monitoring*: A dedicated layer tracks dynamic environmental and social risks, reinforcing strategic planning through real-time, AI-powered assessments of double materiality.

Industry 6.0 redefines the ceramic value chain as a techno-socio-industrial ecosystem, where systemic sustainability and cognitive automation become the pillars of competitive and regulatory alignment.

4.5. From Human-Centric to Systemic Sustainability in Industry 6.0

The transition from Industry 5.0 to Industry 6.0 entails a shift from human-centered innovation to a systemic perspective that integrates technological, social, and environmental dimensions [72,73]. Based on our empirical findings, this transformation can be articulated across four key axes:

1. From Human-Centric to Human–Machine Symbiosis: While Industry 5.0 focused on augmenting human capabilities through collaborative robotics and AI, Industry 6.0 introduces adaptive, co-evolutionary collaboration between humans and machines within real-time digital ecosystems.
2. From Operational Transparency to Cognitive Automation: The emphasis on visibility and control through digital twins in Industry 5.0 evolves into predictive, learning-based systems in Industry 6.0, capable of autonomous decision-making and proactive resource management.
3. From Circular Economy to Systemic Sustainability: Industry 6.0 expands the focus from material recirculation to a broader model of sustainability, integrating environmental stewardship with social resilience and technological innovation.
4. From Fragmented Platforms to the Technosphere: Industry 6.0 introduces a unified digital infrastructure, which the recent literature defines as the Technosphere, that supports end-to-end orchestration of data, processes, and stakeholder interactions.

This evolution is not merely conceptual. Our qualitative evidence reveals that ceramic industry stakeholders are already anticipating this shift. Beyond traditional sustainability goals, they emphasize the integration of AI-powered platforms, co-creation interfaces, and predictive analytics as enablers of systemic resilience. This perspective aligns only partially with existing academic literature, which often remains anchored to circular economy principles. Practitioners, instead, envision a transformation that redefines how value is created and sustained—socially, technologically, and economically. The proposed model (Figure 1) reflects this shift and offers a foundation for future empirical exploration. Translating it into a structured assessment tool would support companies in mapping their Industry 6.0 readiness and identifying gaps in their sustainability strategy.

4.6. Toward a Systemic Value Creation Model in Industry 6.0

Building on the foundational framework proposed in Figure 1, the refined model illustrated in Figure 6 presents an integrated and dynamic configuration of five interdependent pillars.

The updated conceptual model (Figure 6) introduces a sixth pillar (Cognitive Adaptivity), highlighting the role of autonomous learning systems and anticipatory intelligence in Industry 6.0. This addition marks a clear evolution from Industry 5.0, where human-centric collaboration is key but not yet fully integrated with self-adaptive cognitive systems. The six propositions (P1–P6) together define a dynamic framework for systemic, ethical, and resilient value creation. Each pillar contributes to a holistic and ethically grounded system, orchestrated by the Industry 6.0 Paradigm (P6), which serves both as a strategic enabler and a guiding principle.

- *P1—Cognitive Digital Ecosystem:* Evolves from traditional digital transformation to include AI, digital twins, and cognitive automation. It enables real-time optimization of processes and resource flows, anchoring technological intelligence within the broader Industry 6.0 vision.

- *P2—Human–Machine Collaboration:* Extends the human-centric logic of Industry 5.0 by enabling continuous co-evolution between workers and adaptive AI systems. It ensures that technological advancement is aligned with human values, safety, and capabilities.
- *P3—Systemic Sustainability:* Moves beyond circularity by integrating environmental, social, economic, and technological dimensions. It promotes resilience, low-impact design, and inclusive growth across the value chain.
- *P4—Ethics and Governance:* Reinterprets regulatory compliance considering ESG imperatives. It embeds ethical responsibility into innovative processes and supports transparent, accountable governance structures.
- *P5—Holistic Value Ecosystem:* Synthesizes the outcomes of the previous pillars by fostering inclusive, sustainable, and long-term value creation for all stakeholders, from producers to end-users and communities.

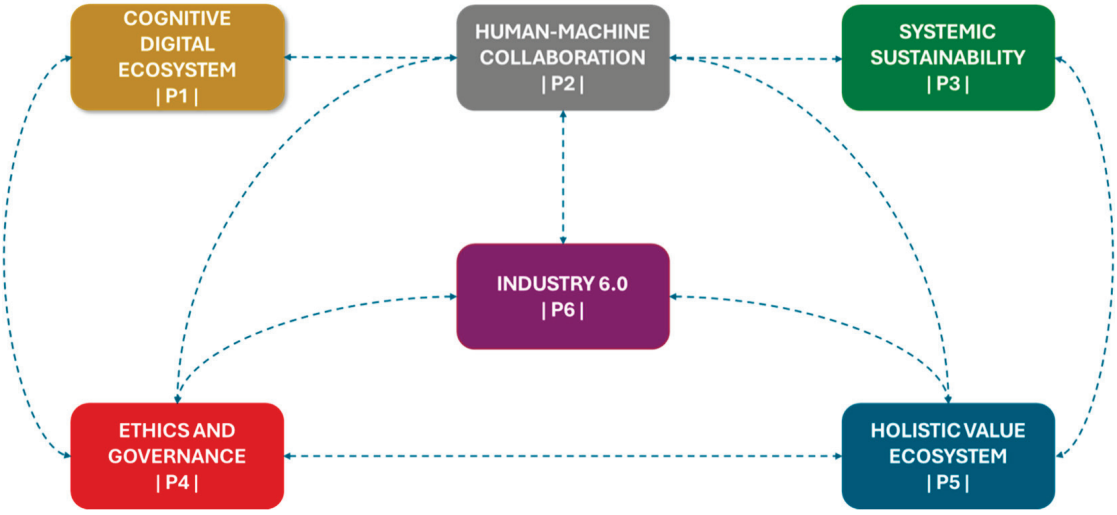


Figure 6. Extended Industry 6.0 model with six pillars, adding Cognitive Adaptivity (P6). As the key differentiator from Industry 5.0.

The Industry 6.0 Paradigm (P6) sits at the center of this model, enabling synergistic feedback loops among the pillars. Rather than acting as a static classification, the model defines a strategic and operational framework for advancing sustainable transformation. It invites future empirical validation to test how these interdependencies manifest in practice and how they can be leveraged to address complexity, uncertainty, and ethical accountability in industrial value chains.

5. Final Remarks

5.1. Discussion of Results

This study shows how the ceramic industry is transitioning from efficiency-driven automation to a systemic integration of AI, sustainability, and human–machine collaboration. The evolution from Industry 3.0 to 6.0 is not linear but cumulative, progressively building an ecosystem where predictive analytics, circularity, and strategic foresight converge. While Industry 4.0 laid the groundwork with IoT and automation, Industry 5.0 introduced personalization and human-centricity. Industry 6.0 amplifies these with cognitive capabilities, embedding sustainability and ethical governance into all value-generating processes. A key theoretical contribution is the articulation of the Technosphere and Systemic Sustainability as interpretive lenses for Industry 6.0. Unlike prior paradigms, the Technosphere conceptualizes industrial value creation as an entangled network of social, technological, and

ecological elements. Systemic Sustainability reframes circular economy logic by integrating digital ethics, real-time risk assessment, and AI-driven responsiveness across the value chain. At the operational level, the results show how AI facilitates predictive maintenance, energy optimization, and cognitive logistics. Stakeholders in the ceramic sector are already embedding these capabilities within digital twins, smart kilns, and advanced quality control systems, validating the centrality of cognitive automation. These transformations are not merely technological upgrades, but strategic enablers for regulatory compliance, resource efficiency, and adaptive B2B collaboration.

The empirical findings can be explicitly mapped to the theoretical propositions developed earlier. For example, the evidence of predictive maintenance systems and IoT adoption validates Proposition 1 (Digital Transformation). Insights on human–AI collaboration and worker safety reinforce Proposition 2 (Human-Centric Innovation). References to circular flows of raw materials and regulatory adaptation by suppliers support Proposition 3 (Resilience and Circularity). The strong role of the CSRD, repeatedly emphasized by both manufacturers and associations, substantiates Proposition 4 (Regulatory Compliance). Finally, the integration of these elements into a systemic framework aligns with Proposition 5 (Holistic Value Creation). Table 3 summarizes these linkages.

Table 3. Summary of key empirical findings and their supported research propositions.

Empirical Evidence (Findings)	Proposition Supported
Predictive maintenance, IoT-based monitoring	P1—Digital Transformation
Human–AI collaboration, safety and customization	P2—Human-Centric Innovation
Circular sourcing, waste reduction practices	P3—Resilience and Circularity
CSRD compliance, ESG reporting mechanisms	P4—Regulatory Compliance
Integrated systemic approach across stakeholders	P5—Holistic Value Creation

Our findings also contribute to the Sustainable Development Goals (SDGs). By demonstrating how AI-driven logistics reduce energy use and emissions, the study aligns with SDG 12 (Responsible Consumption and Production) and SDG 13 (Climate Action). The emphasis on circular flows and cognitive adaptivity supports SDG 9 (Industry, Innovation, and Infrastructure), highlighting how Industry 6.0 can contribute to sustainable industrial ecosystems at both the firm and sectoral levels.

5.2. Addressing Research Questions

The results of the study directly address its research questions (RQs):

- *RQ1* explored how Industry 6.0 technologies integrate with circular economy principles to enhance resilience and value creation in B2B networks. The findings confirm that digital twins, AI, and IoT create interconnected supply systems that optimize performance and comply with demanding ESG frameworks like the CSRD.
- *RQ2* focused on the impact of Industry 6.0 on stakeholder engagement and strategic decision-making in industrial marketing. The research reveals how cognitive platforms and AI-driven ecosystems foster transparency, co-creation, and customization, reshaping traditional B2B interactions into trust-based, value-driven networks.

5.3. Conclusions

This study offers one of the first comprehensive models of Industry 6.0 applied to a resource-intensive manufacturing sector. It advances theory by moving beyond isolated digital or circular strategies and proposing an integrated framework grounded in cognitive ecosystems, systemic sustainability, and ethical value creation. From a managerial perspective, the research suggests that Industry 6.0 technologies are not optional innovations but

strategic imperatives. Cognitive automation, risk monitoring, and human–machine symbiosis enable companies to reconcile regulatory pressures with operational performance. For industries navigating the shift from Industry 4.0 to 6.0, these findings provide a roadmap for adaptive and resilient value chain transformation. It is important to stress that Industry 6.0 remains a largely conceptual paradigm. The findings presented here should therefore be interpreted as indicative trends rather than definitive outcomes. The evidence illustrates emerging trajectories but requires further empirical testing to validate their generalizability beyond the ceramic sector.

Overall, our study not only advances the theoretical debate on Industry 6.0 but also offers practical pathways for aligning industrial transformation with global sustainability agendas. By explicitly linking Industry 6.0 to the SDGs, we highlight its potential as a strategic enabler of long-term competitiveness, resilience, and societal value.

5.4. Limitations and Future Research Directions

As a single-case study in the ceramic industry, the findings require further empirical validation across different manufacturing contexts. Another limitation concerns the composition of the sample. Tile manufacturers dominated the interviews, which, while justified by their pivotal role in the industry, reduced the diversity of perspectives. Future research should aim for a more balanced distribution of stakeholders to capture a richer spectrum of viewpoints across the value chain. For future research, attention should be directed towards cross-sectoral applications of the proposed model and its scalability within small and medium-sized enterprises. Longitudinal studies are essential to examine how cognitive technologies evolve and influence decision-making, sustainability, and competitiveness over time. Furthermore, quantitative validation of the conceptual framework in various manufacturing sectors (such as steel, cement, or automotive) would allow for assessment of its external validity. Comparative analyses across industries and over extended periods would provide deeper insight into the impact of cognitive automation and regulatory compliance on sustainable value creation. Finally, future research should also explore how cognitive adaptivity (P6) can be operationalized and measured in practice, for example, through maturity models or quantitative performance indicators. Cross-sectoral studies in energy-intensive industries such as steel, cement, or chemicals would help assess the scalability of our conceptual model. Longitudinal studies are particularly relevant to capture the evolving role of cognitive systems and their interaction with regulatory frameworks and sustainability outcomes.

Author Contributions: Conceptualization, A.F.-M. and D.S.-B.; methodology, S.O.-M. and M.J.-C.; investigation, A.F.-M. and D.S.-B.; data curation, A.P.F.d.H. and F.E.G.-M.; writing—original draft preparation, D.S.-B. and A.P.F.d.H.; validation, M.J.-C.; supervision, S.O.-M. and F.E.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cheng, C.T.; Wollert, J.; Chen, X.; Fapojuwo, A.O. Guest Editorial Circuits and Systems for Industry X.0 Applications. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2023**, *13*, 457–460. [CrossRef]
2. Appolloni, A.; D'Adamo, I.; Gastaldi, M.; Yazdani, M.; Settembre-Blundo, D. Reflective Backward Analysis to Assess the Operational Performance and Eco-Efficiency of Two Industrial Districts. *Int. J. Product. Perform. Manag.* **2023**, *72*, 1608–1626. [CrossRef]

3. Nasir, N.; Umar, M.; Khan, S.; Zia-ul-haq, H.M.; Yusliza, M.Y. Technological Revolution in Industrial Ecology. In *Energy Transition: Economic, Social and Environmental Dimensions*; Khan, S.A.R., Panait, M., Puime Guillen, F., Raimi, L., Eds.; Springer Nature: Singapore, 2022; pp. 1–28, ISBN 978-981-19-3540-4.
4. Zakoldaev, D.A.; Shukalov, A.V.; Zharinov, I.O. From Industry 3.0 to Industry 4.0: Production Modernization and Creation of Innovative Digital Companies. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *560*, 012206. [CrossRef]
5. Deepti Raj, G.; Prabadevi, B.; Gopal, R. Evolution of Industry 4.0 and Its Fundamental Characteristics. In *Digital Transformation: Industry 4.0 to Society 5.0*; Kumar, A., Sagar, S., Thangamuthu, P., Balamurugan, B., Eds.; Springer Nature: Singapore, 2024; pp. 1–25, ISBN 978-981-99-8118-2.
6. Ghobakhloo, M.; Iranmanesh, M.; Tseng, M.-L.; Grybauskas, A.; Stefanini, A.; Amran, A. Behind the Definition of Industry 5.0: A Systematic Review of Technologies, Principles, Components, and Values. *J. Ind. Prod. Eng.* **2023**, *40*, 432–447. [CrossRef]
7. Shahzadi, G.; Jia, F.; Chen, L.; John, A. AI Adoption in Supply Chain Management: A Systematic Literature Review. *J. Manuf. Technol. Manag.* **2024**, *35*, 1125–1150. [CrossRef]
8. Grabowska, S.; Saniuk, S.; Gajdzik, B. Industry 5.0: Improving Humanization and Sustainability of Industry 4.0. *Scientometrics* **2022**, *127*, 3117–3144. [CrossRef]
9. Caiazzo, B.; Murino, T.; Petrillo, A.; Piccirillo, G.; Santini, S. An IoT-Based and Cloud-Assisted AI-Driven Monitoring Platform for Smart Manufacturing: Design Architecture and Experimental Validation. *J. Manuf. Technol. Manag.* **2023**, *34*, 507–534. [CrossRef]
10. Contini, G.; Peruzzini, M.; Bulgarelli, S.; Bosi, G. Developing Key Performance Indicators for Monitoring Sustainability in the Ceramic Industry: The Role of Digitalization and Industry 4.0 Technologies. *J. Clean. Prod.* **2023**, *414*, 137664. [CrossRef]
11. Chourasia, S.; Tyagi, A.; Pandey, S.M.; Walia, R.S.; Murtaza, Q. Sustainability of Industry 6.0 in Global Perspective: Benefits and Challenges. *MAPAN* **2022**, *37*, 443–452. [CrossRef]
12. Yadav, S.; Rab, S.; Wan, M. Metrology and Sustainability in Industry 6.0: Navigating a New Paradigm. In *Handbook of Quality System, Accreditation and Conformity Assessment*; Springer Nature: Singapore, 2024; pp. 1–31.
13. Dondi, M.; García-Ten, J.; Rambaldi, E.; Zanelli, C.; Vicent-Cabedo, M. Resource Efficiency versus Market Trends in the Ceramic Tile Industry: Effect on the Supply Chain in Italy and Spain. *Resour. Conserv. Recycl.* **2021**, *168*, 105271. [CrossRef]
14. Pagano, A.; Carloni, E.; Galvani, S.; Bocconcelli, R. The Dissemination Mechanisms of Industry 4.0 Knowledge in Traditional Industrial Districts: Evidence from Italy. *Compet. Rev. Int. Bus. J.* **2021**, *31*, 27–53. [CrossRef]
15. Primec, A.; Belak, J. Sustainable CSR: Legal and Managerial Demands of the New EU Legislation (CSRD) for the Future Corporate Governance Practices. *Sustainability* **2022**, *14*, 16648. [CrossRef]
16. Ostojic, S.; Simone, L.; Edler, M.; Traverso, M. How Practically Applicable Are the EU Taxonomy Criteria for Corporates?—An Analysis for the Electrical Industry. *Sustainability* **2024**, *16*, 1575. [CrossRef]
17. Elhusseiny, H.M.; Crispim, J. A Review of Industry 4.0 Maturity Models: Theoretical Comparison in The Smart Manufacturing Sector. *Procedia Comput. Sci.* **2024**, *232*, 1869–1878. [CrossRef]
18. Zamora Iribarren, M.; Garay-Rondero, C.L.; Lemus-Aguilar, I.; Peimbert-García, R.E. A Review of Industry 4.0 Assessment Instruments for Digital Transformation. *Appl. Sci.* **2024**, *14*, 1693. [CrossRef]
19. Kaswan, M.S.; Chaudhary, R.; Garza-Reyes, J.A.; Singh, A. A Review of Industry 5.0: From Key Facets to a Conceptual Implementation Framework. *Int. J. Qual. Reliab. Manag.* **2025**, *42*, 1196–1223. [CrossRef]
20. Rijwani, T.; Kumari, S.; Srinivas, R.; Abhishek, K.; Iyer, G.; Vara, H.; Dubey, S.; Revathi, V.; Gupta, M. Industry 5.0: A Review of Emerging Trends and Transformative Technologies in the next Industrial Revolution. *Int. J. Interact. Manuf.* **2025**, *19*, 667–679. [CrossRef]
21. Nair, A.; Pillai, S.V.; Senthil Kumar, S.A. Towards Emerging Industry 5.0—A Review-Based Framework. *J. Strategy Manag.* **2024**. [CrossRef]
22. Subbiah, P.; Tyagi, A.K.; Mazumdar, B.D. The Future of Manufacturing and Artificial Intelligence Industry 6.0 and Beyond. In *Industry 4.0, Smart Manufacturing, and Industrial Engineering*; CRC Press: Boca Raton, FL, USA, 2024; pp. 347–362.
23. Samuels, A. Examining the Integration of Artificial Intelligence in Supply Chain Management from Industry 4.0 to 6.0: A Systematic Literature Review. *Front. Artif. Intell.* **2025**, *7*, 1477044. [CrossRef]
24. Kumar, U.; Kumar, R.; Singh, M.; Singh, M.; Singh, H.; Singh, R.; Chabra, R. A Critical Review on History of Industrial Revolutions. *AIP Conf. Proc.* **2025**, *3185*, 020097. [CrossRef]
25. Fernández-Miguel, A.; García-Muiña, F.E.; Settembre-Blundo, D.; Tarantino, S.C.; Riccardi, M.P. Exploring Systemic Sustainability in Manufacturing: Geoanthropology's Strategic Lens Shaping Industry 6.0. *Glob. J. Flex. Syst. Manag.* **2024**, *25*, 579–600. [CrossRef]
26. Tyagi, A.K.; Bhatt, P.; Chidambaram, N.; Kumari, S. Artificial Intelligence Empowered Smart Manufacturing for Modern Society. In *Artificial Intelligence-Enabled Digital Twin for Smart Manufacturing*; Wiley: Hoboken, NJ, USA, 2024; pp. 55–83.
27. Sharma, A.; Sim, K.Y.; Chandrasekaran, S. A Comprehensive Review of Challenges Using AI for Smart Manufacturing. In Proceedings of the 2025 17th International Conference on Computer and Automation Engineering (ICCAE), Perth, Australia, 20 March 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 405–413.

28. Ponnusamy, V.; Ekambaram, D.; Zdravkovic, N. Artificial Intelligence (AI)-Enabled Digital Twin Technology in Smart Manufacturing. In *Industry 4.0, Smart Manufacturing, and Industrial Engineering*; CRC Press: Boca Raton, FL, USA, 2024; pp. 248–270.
29. European Commission: Directorate-General for Research and Innovation. *Industry 5.0, A Transformative Vision for Europe—Governing* Systemic Transformations Towards a Sustainable Industry; Publications Office of the European Union: Luxembourg, 2021. Available online: https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/industry-50-transformative-vision-europe_en (accessed on 18 September 2025).
30. European Commission: Directorate-General for Research and Innovation. *ERA Industrial Technologies Roadmap on Human-Centric Research and Innovation for the Manufacturing Sector*; Publications Office of the European Union: Luxembourg, 2024. Available online: https://data.europa.eu/doi/10.2777/0266 (accessed on 18 September 2025).
31. Chui, M.; Hazan, E.; Roberts, R.; Singla, A.; Smaje, K. The Economic Potential of Generative AI. McKinsey. 2023. Available online: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier (accessed on 18 September 2025).
32. Duggal, A.S.; Malik, P.K.; Gehlot, A.; Singh, R.; Gaba, G.S.; Masud, M.; Al-Amri, J.F. A Sequential Roadmap to Industry 6.0: Exploring Future Manufacturing Trends. *IET Commun.* **2022**, *16*, 521–531. [CrossRef]
33. Reddy, C.K.K.; Doss, S.; Bhatia Khan, S.; Alqhatani, A. *Evolution and Advances in Computing Technologies for Industry 6.0: Technology, Practices, and Challenges*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2024; ISBN 9781003503934.
34. Porter, M.E. TECHNOLOGY AND COMPETITIVE ADVANTAGE. *J. Bus. Strat.* **1985**, *5*, 60–78. [CrossRef]
35. Ueda, K.; Takenaka, T.; Fujita, K. Toward Value Co-Creation in Manufacturing and Servicing. *CIRP J. Manuf. Sci. Technol.* **2008**, *1*, 53–58. [CrossRef]
36. Holweg, M.; Helo, P. Defining Value Chain Architectures: Linking Strategic Value Creation to Operational Supply Chain Design. *Int. J. Prod. Econ.* **2014**, *147*, 230–238. [CrossRef]
37. Taşkan, B.; Karatop, B.; Kubat, C. Impacts of Industrial Revolutions on the Enterprise Performance Management: A Literature Review. *J. Bus. Manag.* **2020**, *26*, 79–119. [CrossRef]
38. Kagermann, H.; Wahlster, W.; Helbig, J. Securing the Future of German Manufacturing Industry: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0. *Final. Rep. Ind.* **2013**, *4*. Available online: https://www.din.de/resource/blob/76902/e8cac883f42bf28536e7e8165993f1fd/recommendations-for-implementing-industry-4-0-data.pdf (accessed on 18 September 2025).
39. Lasi, H.; Fettke, P.; Kemper, H.-G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [CrossRef]
40. Kamble, S.S.; Gunasekaran, A.; Parekh, H.; Mani, V.; Belhadi, A.; Sharma, R. Digital Twin for Sustainable Manufacturing Supply Chains: Current Trends, Future Perspectives, and an Implementation Framework. *Technol. Forecast. Soc. Change* **2022**, *176*, 121448. [CrossRef]
41. Nahavandi, S. Industry 5.0—A Human-Centric Solution. *Sustainability* **2019**, *11*, 4371. [CrossRef]
42. Ghobakhloo, M.; Iranmanesh, M.; Mubarak, M.F.; Mubarik, M.; Rejeb, A.; Nilashi, M. Identifying Industry 5.0 Contributions to Sustainable Development: A Strategy Roadmap for Delivering Sustainability Values. *Sustain. Prod. Consum.* **2022**, *33*, 716–737. [CrossRef]
43. Kumar, U.; Kaswan, M.S.; Kumar, R.; Chaudhary, R.; Garza-Reyes, J.A.; Rathi, R.; Joshi, R. A Systematic Review of Industry 5.0 from Main Aspects to the Execution Status. *TQM J.* **2024**, *36*, 1526–1549. [CrossRef]
44. Reddy, M.S.; Reddy, C.K.K.; Hanafiah, M.M. Climate Change Mitigation and Adaptation Strategies Enhanced by Intelligent Systems in Industry 6.0. In *Maintaining a Sustainable World in the Nexus of Environmental Science and AI*; IGI Global: Hershey, PA, USA, 2024; pp. 201–228.
45. Carayannis, E.G.; Posselt, T.; Preissler, S. Toward Industry 6.0 and Society 6.0: The Quintuple Innovation Helix With Embedded AI Modalities as Enabler of Public Interest Technologies Strategic Technology Management and Road-Mapping. *IEEE Trans. Eng. Manag.* **2024**, *71*, 11238–11252. [CrossRef]
46. Movahed, A.B.; Movahed, A.B.; Nozari, H.; Rahmaty, M. Security Criteria in Financial Systems in Industry 6.0. In *Advanced Businesses in Industry 6.0*; IGI Global: Hershey, PA, USA, 2024; pp. 62–74.
47. Olaleye, I.A.; Mokogwu, C.; Olufemi-Phillips, A.Q. Titilope Tosin Adewale Real-Time Inventory Optimization in Dynamic Supply Chains Using Advanced Artificial Intelligence. *Int. J. Manag. Entrep. Res.* **2024**, *6*, 3830–3843. [CrossRef]
48. Gupta, S.; Agarwal, S. Applications of Machine Learning and Artificial Intelligence in Environmental, Social and Governance (ESG) Sector. *SSRN Electron. J.* **2025**. [CrossRef]
49. Singh, J.; Hamid, A.B.A.; Garza-Reyes, J.A. Supply Chain Resilience Strategies and Their Impact on Sustainability: An Investigation from the Automobile Sector. *Supply Chain. Manag. Int. J.* **2023**, *28*, 787–802. [CrossRef]
50. Abideen, A.Z.; Pyeman, J.; Sundram, V.P.K.; Tseng, M.-L.; Sorooshian, S. Leveraging Capabilities of Technology into a Circular Supply Chain to Build Circular Business Models: A State-of-the-Art Systematic Review. *Sustainability* **2021**, *13*, 8997. [CrossRef]
51. Geissdoerfer, M.; Savaget, P.; Bocken, N.M.P.; Hultink, E.J. The Circular Economy—A New Sustainability Paradigm? *J. Clean. Prod.* **2017**, *143*, 757–768. [CrossRef]

52. Naghi, L.E.; Onufreiciuc, R.A.; Stanescu, L.-E.; Hodoș, R.F. Policy and Regulatory Framework on Fighting Financial Crime for Developing Sustainable Economy Models. In *Economic and Financial Crime, Sustainability and Good Governance*; Achim, M.V., Ed.; Springer International Publishing: Cham, Switzerland, 2023; pp. 273–296, ISBN 978-3-031-34082-6.
53. European Union. Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 on Corporate Sustainability Reporting. 2022. Available online: http://data.europa.eu/eli/dir/2022/2464/oj (accessed on 18 September 2025).
54. Hahn, R.; Kühnen, M. Determinants of Sustainability Reporting: A Review of Results, Trends, Theory, and Opportunities in an Expanding Field of Research. *J. Clean. Prod.* **2013**, *59*, 5–21. [CrossRef]
55. Adams, C.A. Sustainability Reporting and Value Creation. *Soc. Environ. Account. J.* **2020**, *40*, 191–197. [CrossRef]
56. Eccles, R.G.; Ioannou, I.; Serafeim, G. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Manag. Sci.* **2014**, *60*, 2835–2857. [CrossRef]
57. Hummel, K.; Jobst, D. An Overview of Corporate Sustainability Reporting Legislation in the European Union. *Account. Eur.* **2024**, *21*, 320–355. [CrossRef]
58. Afzal, B.; Li, X.; Hernández-Lara, A.B. The Innovation Journey and Crossroads of Sustainability, Resilience and Human-Centeredness: A Systematic Literature Review. *Transform. Gov. People Process Policy* **2024**, *18*, 368–383. [CrossRef]
59. Chabane, B.; Komljenovic, D.; Abdul-Nour, G. Converging on Human-Centred Industry, Resilient Processes, and Sustainable Outcomes in Asset Management Frameworks. *Environ. Syst. Decis.* **2023**, *43*, 663–679. [CrossRef]
60. Rajumesh, S. Promoting Sustainable and Human-Centric Industry 5.0: A Thematic Analysis of Emerging Research Topics and Opportunities. *J. Bus. Socio Econ. Dev.* **2024**, *4*, 111–126. [CrossRef]
61. Mohanta, P.R.; Mahanty, B. Assessing Industry 4.0 Implementation Maturity in Manufacturing MSMEs—A Graph Theory and Matrix-Based Approach. *TQM J.* **2024**. [CrossRef]
62. Dierwechter, Y. Methodology: Mixed-Methods Research Design. In *Urban Sustainability through Smart Growth: Intercurrence, Planning, and Geographies of Regional Development Across Greater Seattle*; Dierwechter, Y., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 63–71, ISBN 978-3-319-54448-9.
63. Hunziker, S.; Blankenagel, M. Single Case Research Design. In *Research Design in Business and Management*; Springer Fachmedien: Wiesbaden, Germany, 2024; pp. 141–170.
64. Zucchella, A.; Sanguineti, F.; Contino, F. Collaborations between MNEs and Entrepreneurial Ventures. A Study on Open Innovability in the Energy Sector. *Int. Bus. Rev.* **2024**, *33*, 102228. [CrossRef]
65. Popa, D.; Repanovici, A.; Lupu, D.; Norel, M.; Coman, C. Using Mixed Methods to Understand Teaching and Learning in COVID 19 Times. *Sustainability* **2020**, *12*, 8726. [CrossRef]
66. Prevett, P.S.; Black, L.; Hernandez-Martinez, P.; Pampaka, M.; Williams, J. Integrating Thematic Analysis with Cluster Analysis of Unstructured Interview Datasets: An Evaluative Case Study of an Inquiry into Values and Approaches to Learning Mathematics. *Int. J. Res. Method Educ.* **2021**, *44*, 273–286. [CrossRef]
67. Habibullah, K.M.; Heyn, H.-M.; Gay, G.; Horkoff, J.; Knauss, E.; Borg, M.; Knauss, A.; Sivencrona, H.; Li, J. Requirements Engineering for Automotive Perception Systems: An Interview Study. In *Requirements Engineering: Foundation for Software Quality*; Ferrari, A., Penzenstadler, B., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 189–205.
68. Picot, P.; Guillaume, B. The Controllability of the Technosphere, an Impossible Question. *Anthr. Rev.* **2024**, *11*, 91–109. [CrossRef]
69. Herrmann-Pillath, C. The Case for a New Discipline: Technosphere Science. *Ecol. Econ.* **2018**, *149*, 212–225. [CrossRef]
70. Zimek, M.; Baumgartner, R.J. Systemic Sustainability Assessment: Analyzing Environmental and Social Impacts of Actions on Sustainable Development. *Clean. Prod. Lett.* **2024**, *7*, 100064. [CrossRef]
71. Pascarella, L.; Bednar, P. Systemic Sustainability as Multiple Perspective Analysis. In *Exploring Digital Resilience*; Cuel, R., Ponte, D., Virili, F., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 69–86.
72. Patalas-Maliszewska, J.; Szmołda, M.; Łosyk, H. Integrating artificial intelligence into the supply chain in order to enhance sustainable production—A systematic literature review. *Sustainability* **2024**, *16*, 7110. [CrossRef]
73. Narkhede, G.B.; Pasi, B.N.; Rajhans, N.; Kulkarni, A. Industry 5.0 and sustainable manufacturing: A systematic literature review. *Benchmarking Int. J.* **2025**, *32*, 608–635. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

The Role of Multi-Agent Systems in Realizing Asset Administration Shell Type 3

Lucas Sakurada [1,*], Fernando De la Prieta [2] and Paulo Leitao [1,*]

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal
[2] BISITE Digital Innovation Hub, Edificio Multiusos I+D+i, University of Salamanca, 37007 Salamanca, Spain; fer@usal.es
* Correspondence: lsakurada@ipb.pt (L.S.); pleitao@ipb.pt (P.L.)

Abstract

In the context of Industry 4.0 (I4.0), the Asset Administration Shell (AAS) has been gaining significant attention in recent years. The AAS serves as a standardized digital representation of an asset, encapsulating all relevant information about the asset throughout its lifecycle. Since its introduction in 2015, the past decade has seen considerable progress in developing traditional AAS solutions, namely AAS Type 1 and Type 2. As this initial phase reaches maturity, it becomes essential to shift focus toward AAS Type 3 (proactive), a specific category that extends traditional AAS functionalities by incorporating higher levels of autonomy, intelligence, and collaborative capabilities. However, AAS Type 3 is still in its early stages, lacking formal specifications and comprehensive implementation guidelines. In this context, Multi-Agent Systems (MAS) have been investigated as a means to enhance traditional AAS solutions toward the realization of AAS Type 3, particularly by embedding autonomous, intelligent, and collaborative behaviors. Building on this perspective, this paper explores the role of MAS in realizing AAS Type 3 through a comprehensive analysis of existing agent-based AAS approaches in the literature. Furthermore, this paper proposes a reference model based on common patterns found in the literature to support the development of AAS Type 3 solutions, contributing to the discussion on the formalization of specifications and providing greater clarity on this emerging topic. Finally, to better demonstrate key aspects of the model, some illustrative examples are presented to guide its application and facilitate understanding.

Keywords: Asset Administration Shells; Industry 4.0; Multi-Agent Systems

1. Introduction

Modern factories are becoming increasingly complex and demanding intelligent, connected solutions to support flexible and adaptive operations. However, this evolution introduces significant challenges related to the integration and interoperability of industrial assets. These assets, encompassing a wide range of physical and logical components such as machines, sensors, software, and products, often originate from different vendors and exhibit considerable variation in their communication and functional capabilities. This heterogeneity complicates seamless integration and effective collaboration within increasingly complex industrial ecosystems.

To overcome these challenges, the Asset Administration Shell (AAS) [1,2] has emerged as a fundamental concept in I4.0, providing a standardized digital representation of assets.

The AAS encapsulates all relevant information about an asset throughout its entire lifecycle, enabling consistent identification, access, and management through standardized I4.0-compliant interfaces. By providing a standardized approach for asset description and interaction, the AAS acts as a cornerstone for interoperability, facilitating the integration, communication, and data exchange among heterogeneous devices, machines, and systems in industrial environments.

Since the introduction of AAS in 2015 [1], the past decade has seen considerable progress in developing traditional AAS solutions, namely AAS Type 1 (passive) and Type 2 (reactive) [3]. While these types offer considerable advantages in standardizing asset representation and enabling interoperability, they lack the active behaviors required to support fully autonomous interactions and intelligent decision-making demanded by dynamic I4.0 environments. In this context, AAS Type 3 (proactive) [3] is a distinct category that extends traditional AAS functionalities by incorporating higher levels of autonomy, intelligence, and collaborative capabilities. These features enable AAS Type 3 to autonomously respond to changing production conditions, negotiate with other assets, and optimize operations in real-time, addressing the flexibility, robustness, and reconfigurability demanded by modern industrial systems. Nevertheless, AAS Type 3 remains an underexplored topic in the literature, lacking formal specifications and comprehensive implementation guidelines.

The lack of a detailed specification for AAS Type 3 introduces uncertainty about its structure and implementation. A key question is whether AAS Type 3 requires a uniform, uniquely defined architecture, which may necessitate custom software solutions for its development and deployment, or if it allows for more flexible interpretations. For instance, the literature has increasingly recognized Multi-Agent Systems (MAS) [4] as a promising and suitable approach to support the implementation of AAS Type 3, given their conceptual alignment and inherent support for autonomy, intelligence, and collaborative capabilities [5–7]. Additionally, MAS have been developed and applied for decades in various domains, demonstrating their benefits and effectiveness in enabling decentralized control, adaptive behavior, and dynamic coordination among distributed entities [8,9], which are key requirements for realizing AAS Type 3 solutions.

Given the incipient nature of AAS Type 3, the lack of formal specifications, and the emerging trend of leveraging MAS to support its implementation, this paper explores the role of MAS in realizing AAS Type 3 through a comprehensive analysis of existing agent-based AAS approaches in the literature. Based on the insights gained from this analysis, this paper proposes an agent-based reference model for AAS Type 3 that offers a modular and standardized foundation to support the development of such solutions. This model consolidates and formalizes recurring architectural patterns and design guidelines identified, contributing to the discussion on the formalization of specifications and providing greater clarity on this emerging topic. Moreover, it extends the existing AAS structure [10], facilitating the seamless transition from established Type 1 and Type 2 implementations toward more autonomous and intelligent Type 3 solutions. Finally, to better demonstrate key aspects of the model, some illustrative examples are presented to guide its application and facilitate understanding.

The rest of the paper is organized as follows: Section 2 provides an overview of AAS, AAS Type 3, and MAS concepts. Section 3 explores the role of MAS in realizing AAS Type 3 through a comprehensive analysis of existing agent-based AAS approaches found in the literature. Section 4 proposes an agent-based reference model for realizing AAS Type 3. Finally, Section 5 concludes this paper and presents the potential future work.

2. Background

Several reference architectures have been established to facilitate the implementation of I4.0 [11]. Among them, the Reference Architecture Model Industrie 4.0 (RAMI4.0) [2] and the Industrial Internet Reference Architecture (IIRA) [12] stand out as the two most prominent. Although both offer valuable guidance for the development of I4.0-compliant solutions, and the IIRA is recognized as a complementary perspective to RAMI4.0 [13], this work is exclusively positioned within the RAMI4.0 framework, which formally defines and promotes the AAS as a core enabler of digitization and interoperability in industrial environments. In this context, this section introduces the foundational concepts of the AAS and MAS. These two concepts have gained increasing attention in the literature as complementary approaches for enhancing the intelligence, autonomy, collaboration, and interoperability of I4.0 solutions.

2.1. Asset Administration Shell

The so-called I4.0 component, introduced by the RAMI4.0, encompasses an asset and its corresponding AAS. The AAS is defined as a standardized digital representation of an asset, encapsulating all relevant asset information throughout its lifecycle [10]. As illustrated in Figure 1 (right), the AAS has a well-defined structure comprising a collection of submodels, where each submodel can be viewed as a modular building block that encapsulates specific information about the asset, such as identification, capabilities, technical data, and operational data.

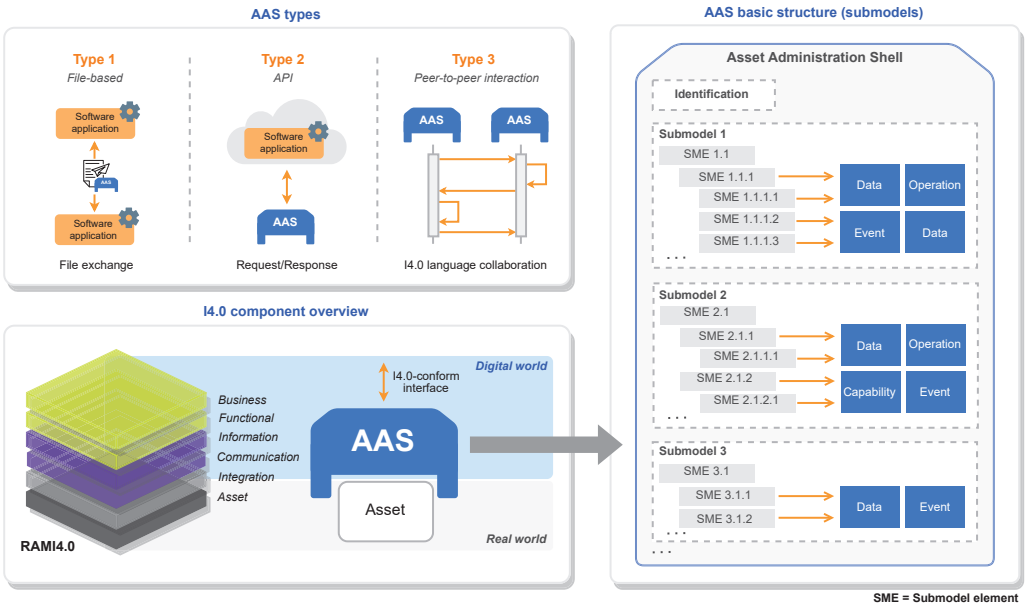


Figure 1. AAS overview (**bottom**), AAS types (**top**), and AAS basic structure (**right**).

Today, the Industrial Digital Twin Association (IDTA) is responsible for standardizing the AAS, providing a set of specifications and submodel templates to support AAS implementations. For instance, *Part 1: Metamodel* specification [10] describes a technology-neutral AAS information metamodel, presenting the main classes and their relationships to develop AASs. On the other hand, *Part 2: Application Programming Interfaces* specification [14] defines APIs for enabling access to the information provided by an AAS. Additionally, submodel templates guide the creation of submodels based on predefined structures. For instance, the *Asset Interfaces Description* [15] submodel template specifies an information model and a common representation for describing the interfaces of an asset service or asset-related service.

Additionally, as shown in Figure 1 (top) and described in Table 1, AASs can be classified according to their interaction pattern to exchange information and degree of autonomy to make decisions [3].

Table 1. Description of AAS types.

AAS Type	Description	Interaction Pattern
Type 1 (passive)	Represents the simplest type of AAS (file-based AAS) that stores the asset information throughout its lifecycle and can be exchanged as a whole in the form of a machine- and human-readable file, e.g., XML, JSON, and AASX	File exchange (manual)
Type 2 (reactive)	Acts as an API that reacts to external requests but lacks the capability to take initiatives and decisions. Such API enables the online access to asset information and can be specified in a technology-neutral way	Request/response (semi-autonomous)
Type 3 (proactive)	Represents a more advanced and extended form of AAS, acting as a decision-making entity that autonomously interacts with other AASs to exchange information and collaborate (e.g., through negotiation)	Peer-to-peer interaction (fully autonomous)

The central idea behind the concept is that each asset in the industry is represented by its own AAS, transforming it into an I4.0 component. In this context, since all AASs adhere to a common standardized structure and expose I4.0-compliant interfaces, they can seamlessly provide and exchange information throughout the asset's entire lifecycle. This standardized communication enables interoperability and real-time data sharing among heterogeneous assets, regardless of the manufacturer or domain. Consequently, through their AASs, assets are expected to collaborate autonomously, coordinate processes, and support decentralized decision-making within industrial environments. However, this vision of autonomous, intelligent, and collaborative behavior is more closely aligned with AAS Type 3, which remains relatively underexplored compared to the other types, namely AAS Type 1 and Type 2, and currently lacks dedicated specifications and comprehensive implementation guidelines.

Finally, a recurring question in the community concerns the relationship between the AAS and the Digital Twin (DT) since both are defined as digital representations of assets. Given the growing popularity of both concepts, several studies [16–18] have investigated their interrelation and distinctions. These studies reveal that the relationship between AAS and DT is subject to varying interpretations. Some sources treat the AAS as a synonym for or an implementation approach to DTs, while others regard it as a foundational information model that enables the creation and interoperability of DTs. Moreover, the suitability of AAS as a representation of a DT is seen to increase along the AAS types, from Type 1 to Type 3. In particular, AAS Type 3 is considered to have the strongest alignment with the DT concept, potentially serving as a near-complete implementation technology for DT, as discussed in [18].

It is important to emphasize that this work does not aim to engage in broader conceptual debates. Therefore, it specifically focuses on the AAS, and in particular AAS Type 3, as a standardized framework for advancing digitization and interoperability in I4.0 environments, aligned with the principles initially proposed by Plattform Industrie 4.0 and IDTA.

2.2. Asset Administration Shell Type 3 Definition and Requirements

As previously discussed, AAS Type 3 is still in its early stages. In this context, this section aims to examine and clarify the concept of AAS Type 3 by analyzing how it is defined in official documents, e.g., those provided by the Plattform Industrie 4.0 and IDTA, along with some definitions from selected academic research. The objective is to offer a

clearer understanding of its key characteristics and intended functionalities. Based on this analysis, a general definition and the requirements for AAS Type 3 are established.

2.2.1. Definition

As summarized in Table 2, the definitions of AAS Type 3 presented in official documents, namely [3,19–21], as well as in selected research works, are brief, generic, and remain limited in scope, evidencing the early stage of conceptual development. However, they consistently emphasize several common and core aspects: the ability to interact and collaborate through the I4.0 language (VDI/VDE 2193) and the integration of autonomy and intelligence to support decision-making processes. Additional requirements, as well as concrete guidelines for the design and implementation of such components, are still in the early stages of development, particularly in official documents, which only highlight the importance of adopting the I4.0 language for communication.

Table 2. Summary of AAS Type 3 definitions obtained from official documents and research works.

Ref.	AAS Type 3 (Proactive AAS) Definition
[3] *	"(...) can participate in protocol-based interactions, as defined, for example, in the VDI/VDE 2193 standard for the bidding process (...) The goal is to design decentralized processes that rely on a certain degree of autonomy or decision-making capability of the AASs"
[19] *	"Interactions which need peer-to-peer interactions between AAS of I4.0 components. The exchange of messages may be structured according to a grammar of an Industrie 4.0 language or may be implemented by using standardized APIs. Type 3 AAS interactions enable I4.0 components to become pro-active components which can support system integration and system interaction in a peer-to-peer environment between I4.0 components, e.g., to enable AAS assisted plug & play scenarios between I4.0 components"
[20] *	"An AAS Server Application with a corresponding interface for the I4.0 language is at the same time a client and a server and is a so called proactive AAS"
[21] *	"Type 3 is the so called "Industrie 4.0 Language" with active AAS, which is similar to agents"
[22]	"AASs are able to interact autonomously with each other via a standardized interface with a common syntax and semantics base, thus realizing peer-to-peer AAS interaction. (...) AASs can understand and cooperate with each other. (...) An AAS with such autonomy is called a proactive one"
[23]	"(...) proactive AAS can implement intelligence. The proactive AAS, that is not fully specified yet, can realize semi- or full-autonomous interactions which enables order driven production"
[24]	"AASs with the ability to interact as equals among themselves, offering and requesting services when necessary to meet their goals"
[25]	"(...) can take decisions and participate in protocol-based interactions"
[26]	"(...) proactive AAS also provides CRUD (create, read, update, delete) capabilities but can additionally communicate and bid with other AAS using the I4.0 language"

* Plattform Industrie 4.0 or IDTA official document.

Based on these definitions, this work proposes the following definition for AAS Type 3: *AAS Type 3 represents the next evolutionary step of the AAS concept, extending AAS Type 1 and Type 2 solutions. AAS Type 3 not only serves as a key enabler of interoperability in I4.0 environments but also introduces a higher degree of autonomy, intelligence, and collaborative capabilities. It acts as a decision-making entity capable of autonomously interacting with other AAS Type 3 instances to exchange information and collaborate, e.g., through negotiation. Furthermore, it possesses the intelligence to self-manage its associated asset, leveraging AI techniques to support the decision-making process and adaptive behavior throughout the asset's lifecycle.*

2.2.2. Requirements

From the available definitions, it is possible to identify key requirements/characteristics for AAS Type 3, namely autonomy, intelligence, and collaboration. Although this set of requirements is still limited, these elements provide a useful basis for recognizing and

classifying an AAS solution as Type 3, serving as a first step toward the development of more comprehensive design guidelines and implementation strategies. In this context, and to provide a clearer understanding of these requirements, this work defines them as follows:

- Autonomy: "the ability to make your own decisions without being controlled by anyone else" (Cambridge Dictionary). Autonomy does not mean absolute freedom to decide arbitrarily but rather the capability to make decisions respecting predefined criteria and aligned with business purposes.
- Intelligence: "the ability to learn and understand things quickly and easily" (Cambridge Dictionary). Intelligence can have different levels, e.g., from basic rule-based mechanisms to advanced AI techniques.
- Collaboration: "involving two or more people or organizations working together for a particular purpose" (Cambridge Dictionary). In this context, collaboration refers to AASs working together for a particular purpose. Unlike simple communication, collaboration goes beyond the exchange of information and comprises a deeper, interdependent process where all participants play crucial roles in achieving a common goal.

2.3. Multi-Agent Systems

As previously mentioned, the definitions and requirements associated with AAS Type 3 closely align with the fundamental principles of MAS, particularly regarding autonomy, intelligence, and collaborative behavior. In this context, this section introduces the concepts of agents, industrial agents, and MAS, highlighting the characteristics and functionalities that make MAS a suitable and promising approach for implementing AAS Type 3 solutions.

The development of agent-based applications has gained significant attention due to the distributed and autonomous nature of these software entities. Although no universally accepted definition exists, and the term remains a subject of ongoing debate and controversy, a classic definition of agent is as follows: "An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives" [4]. Moreover, the agents possess several characteristics:

- Encapsulation: Agents represent physical or logical objects within an environment.
- Autonomy: Agents operate without direct human or external intervention and have control over their actions and internal state.
- Learning ability: Agents can learn from their experiences.
- Reactivity: Agents perceive their environment and respond in a timely manner to changes in order to fulfill their design objectives.
- Proactiveness: Agents exhibit goal-directed behavior by taking the initiative to achieve their design objectives.
- Social ability: Agents interact with other agents (and possibly humans) to achieve their design objectives.

Additionally, agents can be specifically designed to meet the I4.0 requirements, known as industrial agents, a specialization of traditional software agents, which inherit their characteristics and are enhanced with additional capabilities to comply with several industrial requirements, namely hardware integration, reliability, fault-tolerance, scalability, industrial standard compliance, quality assurance, resilience, manageability, and maintainability [9].

Generally, industrial agents follow a two-layered approach (see Figure 2) comprising high-level control (HLC) and low-level control (LLC) [9,27]. HLC is responsible for providing intelligence and adaptation features to support decision-making processes, e.g., through

monitoring, diagnosis, prediction, and optimization functions. However, it is incapable of delivering real-time performance. On the other hand, LLC is responsible for executing low-level control tasks that guarantee real-time constraints, often through PLCs or industrial PCs running control programs compliant with IEC 61131, IEC 61499, or similar standards. This separation allows HLC to handle complex, high-level decision-making while LLC maintains stable, real-time operations.

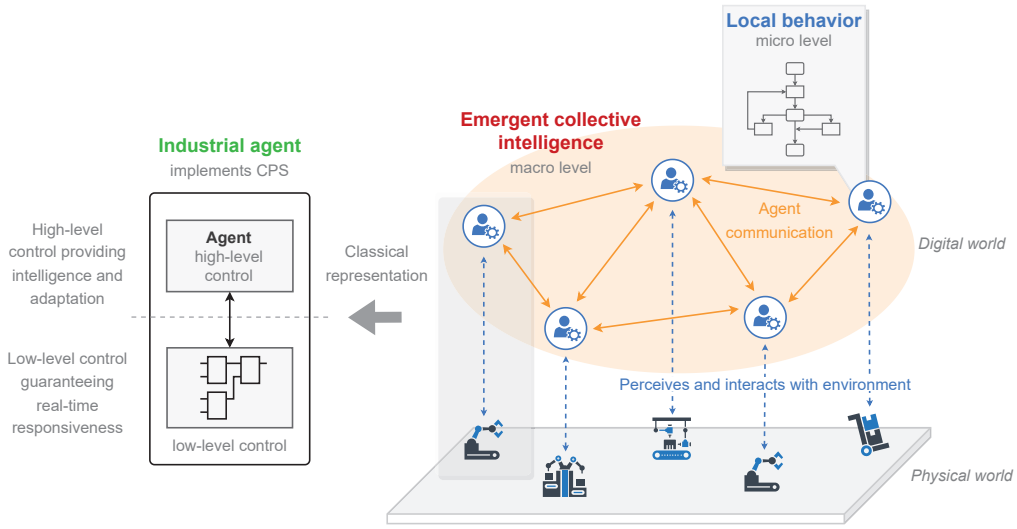


Figure 2. Multi-Agent Systems overview.

Addressing complex real-world problems often requires the combined and cooperative efforts of multiple agents, as a single agent has limited capabilities. As illustrated in Figure 2, a MAS can be described as a society of intelligent, autonomous, and cooperative agents representing physical or logical objects of a system. In contrast to traditional centralized control strategies, MAS offers a decentralized approach to designing large-scale, complex systems by decentralizing the system's control and distributing it to autonomous and cooperative agents rather than being concentrated in a single authority. The overall functionality of the system emerges through interactions among these individual agents, e.g., through collaboration, negotiation, and delegation of responsibilities, where each agent contributes with unique knowledge and skills toward achieving both individual and shared objectives. Furthermore, MAS enables the decomposition of complex tasks into manageable subtasks, each assigned to a capable agent, ensuring that even intricate challenges can be effectively addressed in a decentralized manner [4,8].

Beyond their theoretical foundations, several surveys on MAS (e.g., [8,9,28], to name a few) have highlighted the practical benefits of applying MAS in industrial environments, particularly offering alternatives to overcome the typical problems presented by the traditional centralized control approaches, which are not able to address the flexibility, robustness, and reconfigurability imposed on the current industrial systems.

3. AAS Type 3 Approaches Based on Multi-Agent Systems

As previously discussed, given the advantages provided by MAS, along with their shared characteristics and conceptual alignment with AAS Type 3, particularly regarding autonomy, intelligence, and collaborative capabilities, the current literature has increasingly considered MAS as a suitable approach for enhancing AAS solutions, often referred to as agent-based AAS solutions. Building on this perspective, this section examines the role of MAS in existing agent-based AAS approaches and explores their potential to support the realization of AAS Type 3. Moreover, it is important to highlight that a few other studies are also investigating AAS Type 3, namely [23,25]. However, due to the current trend in

the literature of using MAS to enhance AAS capabilities, this work focuses exclusively on agent-based AAS approaches.

To identify existing agent-based AAS approaches in the literature, a bibliographic search was conducted in the Scopus database using the keywords "Asset Administration Shell" and "Multi-Agent Systems", along with variations of these terms, combined using the AND operator. The search was restricted to publications from 2015 to 2025 (first semester), written in English, and classified as journal articles, conference papers, or book chapters. Additionally, only documents in which the search terms appeared in the title, abstract, or keyword fields were considered.

From the initial set of 35 documents retrieved, a screening process was carried out to identify the most relevant studies for in-depth analysis. Several factors may account for the relatively low number of documents: (i) the AAS concept is still recent, having been formally introduced only in 2015; (ii) much of the research over the past decade has focused on consolidating foundational concepts, particularly related to AAS Types 1 and 2; (iii) overall, the current literature on AAS remains limited, with only 503 documents identified through a broader Scopus search using the keyword "Asset Administration Shell".

Following the screening process, Table 3 summarizes the reviewed research works, offering valuable insights into how MAS have been integrated with AAS to enhance their capabilities. These studies cover a wide range of application domains where agent-based AASs interact to support autonomous, collaborative, and intelligent processes in industrial environments. The identified applications include smart manufacturing, dynamic production replanning, shared production, control, scheduling, human collaboration, diagnosis, and fault detection. This diversity illustrates the versatility of agent-based AASs, showing how they can be effectively adapted to meet the specific demands and operational contexts of various industrial scenarios.

Table 3. Summary of research works that combine MAS and AAS.

Ref.	Application Domain	Solution Description
[26]	Smart manufacturing (prototypical implementation in a demonstrator)	- Discusses that the asset information described in the AAS may be used as a standardized agent's knowledge representation, i.e., the agents get all the relevant information to make decisions from AAS submodels - Assets: demonstrator (a stack, a crane, a stamp, and a conveyor with ramps and pushers)
[29,30]	Smart manufacturing (prototypical implementation in a demonstrator)	- Proposes a design pattern for implementing AASs based on industrial agents. The agents are responsible for decision-making capabilities and integration with physical assets - Assets: product and industrial robot
[31]	Smart manufacturing (simulated experiments)	- Presents a method for developing AASs using the MAS as a key enabler to distribute intelligence and offer collaborative functionalities - Assets: assets simulated in a virtual environment
[32]	Dynamic replanning in production environments (prototypical implementation in a demonstrator)	- Concentrates on modeling AAS submodels for resource and product agents to use in dynamic replanning applications - Assets: demonstrator (a module for transportation, assembly, flashing of information on the product, and two quality control modules)
[33]	Human–robot shared assembly task (prototypical implementation in a demonstrator)	- Proposes a framework that combines MAS with AASs to enable the human-centered production paradigm. The focus is on the active involvement of operators in the industrial collaboration process with other assets - Assets: human, robot, and camera
[34]	Plan and control in industrial automation systems (prototypical implementation in a demonstrator)	- Uses LLM agents to interpret complex information, generate insights, and support decision-making in industrial automation systems. The information is primarily sourced from AASs, which provide descriptive details about a specific asset - Assets: transport robots and painting stations

Table 3. *Cont.*

Ref.	Application Domain	Solution Description
[35]	Smart manufacturing (prototypical implementation in a demonstrator)	- Presents a methodology positioned with respect to the RAMI4.0 layers, which provides guidelines for integrating AASs and MAS - Assets: small-scale production system (robot, punching machines, and indexed line machines)
[36]	Shared production in manufacturing (prototypical implementation in a demonstrator)	- Introduces a MAS architecture designed to enable shared production. The proposed architecture is grounded in holonic principles, and the AASs serve as the self-description mechanism for these agents/holons, providing the necessary information for their operation, interaction, and integration - Assets: robot module, assembly module, 3D printer module, connector module, conveyor module, and quality module
[37]	Production scheduling (deployed in an industrial pilot coming from the bicycle production industry)	- Presents an approach in which agents operate in a decentralized manner, interacting with each other to solve scheduling problems, while the AAS provides production- and agent-related information (e.g., agent identity, description, skills, and configurations) - Assets: production lines (e.g., painting and assembly lines)
[38]	Diagnosis and fault detection in production environments (prototypical implementation in a demonstrator)	- Integrates MAS and AASs to perform fault detection and diagnosis. All asset- and fault-related information is modeled within AAS submodels, while communication between system components is enabled through agent-based interaction using the I4.0 language - Assets: production modules

Overall, there is a general consensus that submodels serve as standardized knowledge representations or information sources for agents. These submodels capture all asset-related information in both human- and machine-readable formats, which agents can leverage to support decision-making processes and to construct or enrich their internal knowledge bases. In this context, several standardized submodels are available for this purpose, e.g., *Control Component Instance* [39], *Asset Interfaces Description* [15], and *Time-Series Data* [40], which can be used to describe the skills provided by an asset (e.g., drilling, assembly, and transport), the methods for interfacing with it (e.g., host address, port, connection state, communication protocol, endpoints, data formats, and input/output parameters), and the mechanisms for managing, accessing, and storing operational data. Additionally, other submodels can be defined to address specific application needs and to represent agent-specific information, such as descriptions of agent behaviors, decision-making strategies, or interaction protocols, providing greater transparency and understanding of agent capabilities.

The standardized structure of submodels enables agents to interpret and process them using suitable programming logic, promoting reusability, scalability, and consistency across heterogeneous assets and application domains. Moreover, agents that are equipped with Large Language Models (LLMs) could enhance their learning and interpretation of submodels. This is possible due to LLMs' ability to process both structured and unstructured data, generalize across different formats, and adapt to varied contexts, ultimately improving the system's flexibility and responsiveness.

The review also reveals that communication capabilities stand out as one of the most emphasized and valuable features of agents. This natural and inherent characteristic of agents positions them as a powerful enabler for achieving the decentralized and collaborative interactions (particularly through negotiation strategies) envisioned in AAS Type 3 implementations.

As shown in Figure 3, in a negotiation-based production scenario, information about capabilities (capability: an abstract description of a function or operation that an asset can perform, which results in tangible real-world effects, such as drilling. This description includes specific properties and constraints that define the scope and limitations of the

operation [41]) and skills (skill: the concrete implementation of a function provided by an asset, as defined by a capability. A skill enables the practical execution of a capability and typically involves parameters, including inputs and outputs, that allow for the control and monitoring of the skill's execution. Each skill must be accessible through at least one interface, which provides interaction with the skill's state machine, as well as its parameters [41]) is organized into submodels, which the agents utilize to manage their respective assets. As a result, agents representing resources possess information of the capabilities their assets can provide and the skills to execute these capabilities. On the other hand, agents representing products possess information of their required capabilities structured in a submodel that outlines the steps and requirements for production based on the required capabilities. This information allows products and resources to collaborate seamlessly to complete the production process, where resources may act as service providers by offering their capabilities as services, while products act as service requesters, seeking services aligned with their production requirements.

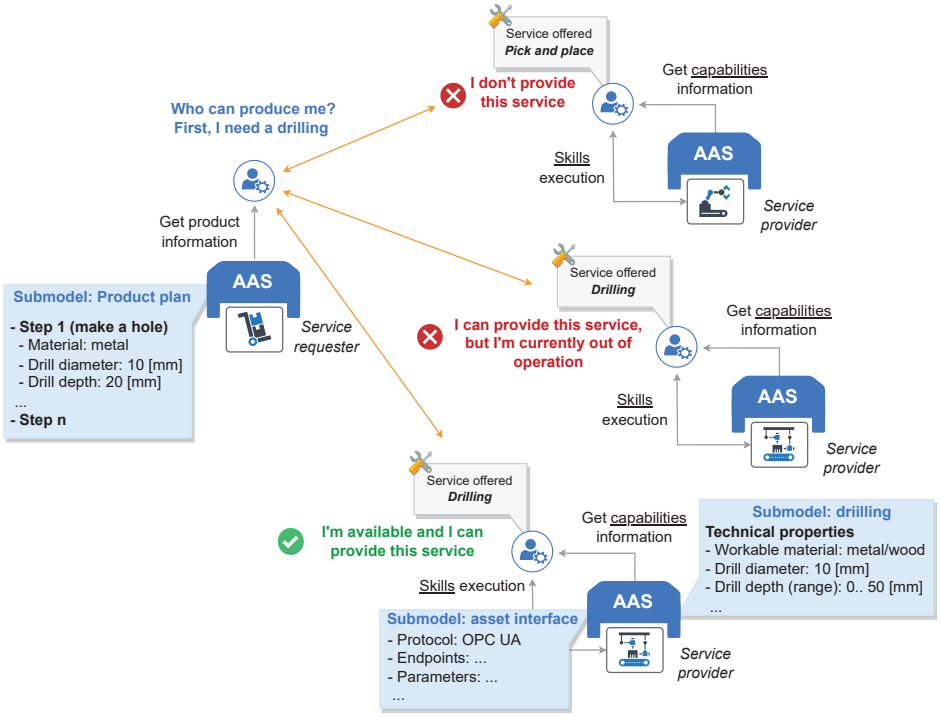


Figure 3. Conceptual view of a negotiation-based production scenario using agent-based AAS.

Moreover, such communication capabilities enabled by agents could open up new possibilities by transcending traditional organizational boundaries. This advancement allows the creation of shared production environments where multiple companies can collaborate in real-time. Through secure and standardized communication protocols, agents representing assets from different organizations can negotiate, allocate resources, and coordinate production tasks seamlessly.

Another important aspect concerns the interface with physical assets for data collection or control adaptation. While the AAS provides I4.0-compliant interfaces to enable seamless data exchange at the information level, it is not inherently designed to interface directly with physical assets. Instead, this connection requires customized implementation tailored to the specific characteristics, communication protocols, and constraints of each asset. In this sense, agents (or well-known industrial agents) are suitable for this purpose, as they are primarily designed to interact with both the physical and digital layers. They can serve as intelligent intermediaries that interpret and process submodel data and translate high-

level commands into asset-specific control actions, thus enabling flexible and autonomous monitoring and control of physical assets.

Although existing agent-based AAS approaches vary from simple rule-based logic to more advanced AI-based behaviors, most current implementations tend to rely primarily on the former, with limited exploration of sophisticated AI techniques. However, the broader agent literature highlights the potential for agents to act as "model consumers" of AI techniques, transitioning toward the so-called AI agents. This includes the integration of advanced learning algorithms, reasoning capabilities, and adaptive behaviors, which could significantly enhance the intelligence and autonomy of AAS-based solutions.

Additionally, some research efforts also explore the use of agents for service orchestration and the adoption of recursive architecture inspired by holonic systems (see Figure 4). In this context, each agent-based AAS encapsulates the capabilities of its associated assets and exposes them as services, enabling the dynamic composition and orchestration of services to realize higher-level services aligned with production goals.

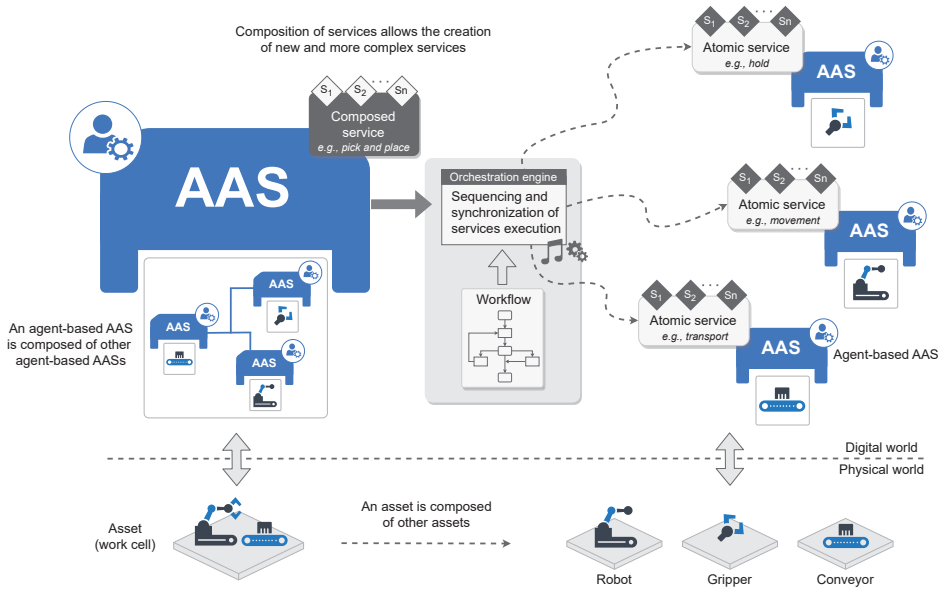


Figure 4. Service orchestration of holonic agent-based AAS.

For example, consider an illustrative example of a work cell comprising a robot, a gripper, and a conveyor. Individually, their respective agent-based AAS expose atomic services such as movement, hold, and transport and are responsible for managing their own services. In this context, the agent-based AAS associated with the work cell contains an orchestration engine to coordinate the execution of these services, relying on effective communication and negotiation between agents to ensure proper sequencing and synchronization. Through this engine, the work cell agent can compose atomic services into a higher-level service, such as pick and place. Extending this concept further, a station agent can orchestrate multiple work cell agents, which in turn coordinate their assets, enabling the composition of complex production tasks to be executed efficiently and flexibly.

Finally, the agent-based AAS approach is not limited to typical physical assets such as machines and production units but is also being applied to represent human operators. In this context, agent-based AASs can be specifically designed to encapsulate operator-related information, such as skills, preferences, fatigue levels, and availability. By leveraging these data, agents can dynamically allocate tasks in a manner that aligns with the operator's current capabilities and conditions, thereby enhancing task performance, ensuring quality of work, and supporting overall human well-being.

In summary, the integration of MAS with AASs represents a promising pathway toward the realization of AAS Type 3, enabling autonomous, intelligent, and collaborative behaviors in industrial environments. The reviewed literature highlights a range of innovative concepts, applications, and use cases, including decentralized negotiation, dynamic service orchestration, and human-centric approaches. Although real-world deployments are currently lacking, simulation and prototype results demonstrate significant potential for enhancing interoperability, flexibility, scalability, responsiveness, and intelligence in industrial systems. Furthermore, most studies are focused on addressing a particular use case and do not provide clear guidance for generalizing the approach or developing comprehensive, reusable solutions. Finally, given that this is a relatively new and evolving field, further research and development are essential to consolidate its theoretical and conceptual foundations before transitioning from the state of the art to practical implementation in industrial environments.

4. Agent-Based Reference Model for Realizing AAS Type 3

Based on the findings of the previous section, there is both theoretical and practical evidence supporting the investigation of MAS as a foundation for realizing AAS Type 3. In this context, and inspired by these insights, a reference model is proposed.

The proposed reference model aims to consolidate and formalize recurring architectural patterns and design guidelines identified by analyzing agent-based AAS approaches from the literature. Most of the approaches reviewed converge around similar solutions and principles. However, they are often presented in isolated or case-specific contexts. This convergence indicates a shared underlying structure that, while implicitly recognized, lacks an explicit and unified representation. The proposed model captures these commonalities, providing a modular and reusable foundation that supports both conceptual clarity and the practical implementation of AAS Type 3 solutions. Moreover, the model serves not only as a general guideline for developing AAS Type 3 but also as a template that offers a high-level structure from which more detailed and application-specific architectures and specifications can be derived.

4.1. Overview

To develop the proposed reference model, a three-step methodology was adopted. First, a literature review was conducted to identify existing approaches integrating AAS and MAS, as discussed in the previous section. Second, these approaches were analyzed to extract common architectural patterns and design guidelines. Third, based on the most frequently observed principles and structures, a generic and reusable reference model was synthesized to guide the development of agent-based AAS solutions, with particular emphasis on supporting AAS Type 3 capabilities.

Additionally, the proposed model (see Figure 5) is designed as a modular extension of the traditional AAS structure [10]. Rather than replacing existing AAS Type 1 or Type 2 implementations, the model builds upon them by introducing an intelligent layer that enables autonomy, collaboration, and adaptive behavior. This modularity allows existing AAS-based solutions to be incrementally extended toward AAS Type 3 capabilities, enhancing their functionality without requiring a complete architectural redesign.

As shown in Figure 5, the defined reference model comprises two complementary parts: the information part and the intelligent part. The information part consists of a structured collection of submodels, which encapsulate detailed asset-specific information. These submodels provide a standardized knowledge base and serve as a comprehensive information source that supports agents in their decision-making processes and effective collaboration within industrial environments. On the other hand, the intelligent part is

conceptualized as an agent that embodies the intelligent capabilities required for AAS Type 3. This agent is responsible for interpreting and processing the information stored in the submodels, making decisions, and executing appropriate actions. Moreover, it facilitates collaboration by communicating and negotiating with other agents, enabling decentralized coordination and adaptive behavior. In summary, the intelligent part goes beyond generic AAS requirements by integrating autonomous decision-making and agent-based collaboration capabilities, which are not formalized in existing AAS specifications but are essential for realizing the full vision of AAS Type 3.

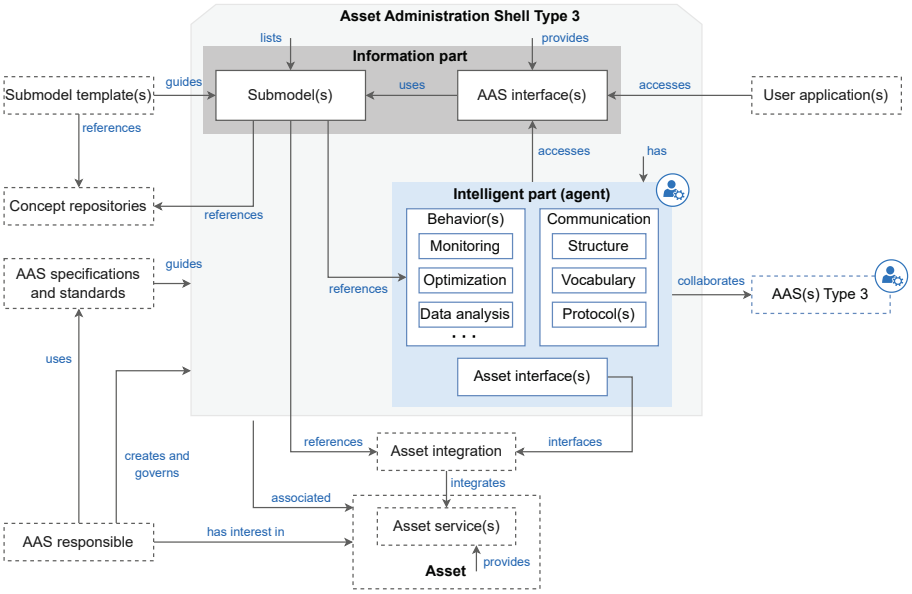


Figure 5. Proposed reference model for AAS Type 3 using Multi-Agent Systems.

As part of the model, a set of design guidelines have been defined to characterize the development of AAS Type 3. These guidelines serve as design principles and requirements, ensuring that any developed solution aligns with the conceptual and technical expectations of AAS Type 3. These guidelines are as follows:

- AAS Type 3 has an associated asset (individual or composite).
- AAS Type 3 provides one or more I4.0-conform interface(s).
- AAS Type 3 lists one or more submodel(s).
- AAS Type 3 uses information from submodels to support decision-making processes, enabling autonomous behavior and intelligent responses.
- AAS Type 3 collaborates with other AAS(s) Type 3 (e.g., through negotiation)
- AAS Type 3 has one or more intelligent behavior(s), e.g., based on AI techniques.
- AAS Type 3 interfaces with the asset, e.g., to collect operational data or adapt control.
- The information and intelligent parts can be deployed together or separately, depending on the system architecture and deployment requirements.
- Submodels reference the asset services provided by an asset via an asset integration and other related services (e.g., based on AI techniques).
- The user application(s) accesses the information of the AAS Type 3 via IT interfaces.
- The submodel template guides the creation of a standardized submodel.
- The submodel template may reference concept dictionaries and ontologies.
- Concept dictionaries and ontologies define the common vocabulary.
- Existing AAS Type 1 and Type 2 specifications and standards guide the development of the information part of AAS Type 3.
- The AAS responsible creates and governs the AAS Type 3.

Although the holonic concept identified in some existing works (see Figure 4) is not part of the proposed reference model, the intelligent part of the model is designed to be flexible enough to support such holonic orchestration mechanisms when needed. However, implementing such approaches would require additional specifications, such as service orchestration and composition mechanisms, which are not fully defined in the current version and are considered directions for future work.

4.2. Alignment with RAMI4.0

While RAMI4.0 provides a valuable framework for developing I4.0-compliant solutions, its high level of abstraction and generality may pose challenges when it comes to implementing autonomous and intelligent behaviors within AAS instances. To address these challenges, the reference model proposed in this work incorporates principles from MAS, enabling a structured approach to designing AAS Type 3. By introducing a separation between the information and intelligent parts, the model provides clearer guidance for integrating autonomy and proactive behavior in a way that remains compatible and aligned with the RAMI4.0.

As previously mentioned, the AAS reflects the asset in the digital world and is conceptually aligned with the upper five layers of RAMI4.0, namely *integration*, *communication*, *information*, *functional*, and *business*. However, the main AAS specifications (Part 1 [10] and Part 2 [14]) primarily address the *communication* and *information* layers. While these existing specifications provide a solid foundation for organizing and exchanging asset information, the *integration*, *functional*, and *business* layers are not explicitly defined.

The *integration*, *functional*, and *business* layers of RAMI4.0 can theoretically be implemented using various approaches/technologies. However, it is crucial to select one that also offers the intelligence, autonomy, and collaboration capabilities envisioned for AAS Type 3. In this context, agents are particularly well-suited to address these requirements due to their unique characteristics, as previously discussed. Agents bring the essential capabilities of autonomy, decision-making, and collaboration, which are critical for enabling AAS to act as a self-managing, intelligent component.

In this context, the *integration* layer can be enhanced through industrial agents, which are typically designed to interface with assets to collect data or to adapt control. The *functional* layer can be improved by incorporating agents that utilize AI techniques for tasks such as monitoring, diagnosis, optimization, and predictive analysis. Additionally, the *business* layer can benefit from agents' ability to make autonomous decisions, optimizing processes like resource allocation, predictive maintenance scheduling, and fault management in alignment with business strategies. Additionally, agents can enhance the *communication* layer by facilitating autonomous, decentralized communication through strategies like collaboration and negotiation. In summary, the map of the proposed model to the RAMI4.0 layers is described in Table 4.

In this context, the benefits of the proposed model are reflected across multiple layers of RAMI4.0, particularly *integration*, *communication*, *functional*, and *business*. It enables the seamless integration of assets into I4.0, supports decentralized communication through strategies such as collaboration and negotiation, empowers assets to perform autonomous functions based on AI techniques, and improves decision-making processes aligned with business objectives.

Since the information part is already well supported by formal specifications and standards provided by the IDTA (e.g., through the metamodel [10] and standardized API interfaces [14], which agents can access via standardized protocols), the following sections discuss how the agent (intelligent part) can be designed in compliance with existing standards and best practices to support broader industrial adoption and enable AAS

Type 3 capabilities. In this sense, the following sections address key features highlighted in Figure 5 (blue box) and Table 4, including interfacing with the physical asset for data collection and control adaptation, communication between AAS Type 3 instances, and strategies for developing AI-based behaviors.

Table 4. Relationship of the proposed model to the layers of RAMI4.0.

RAMI4.0 Layer	Relation to Proposed Model
Asset	Represents the asset, i.e., every physical or logical object that has value for a company
Integration (*intelligent part*)	Industrial agents interface with the asset to collect data or adapt control processes (supported by the IEEE 2660.1 standard [27])
Communication (*information and intelligent part*)	Data exchange follows a client–server schema (supported by AAS Part 2 specification [14]). Additionally, agents enable decentralized communication through interaction strategies such as collaboration and negotiation (supported by FIPA specifications and VDI/VDE 2193 standards [42,43])
Information (*information part*)	Data collected from the asset, including sensor readings and other relevant information, are stored in a standardized and structured manner using AAS submodels (supported by AAS Part 1 specification [10])
Functional (*intelligent part*)	Agents can autonomously execute adaptive functions (e.g., based on AI techniques) for tasks such as monitoring, diagnosis, optimization, and predictive analysis
Business (*intelligent part*)	Agents facilitate self-optimizing, decentralized decision-making aligned with organizational business objectives to enhance operational efficiency

4.3. Recommended Practices for Interfacing with Assets Supported by IEEE 2660.1 Standard

The integration with the asset is a critical aspect of AAS solutions. While the AAS may describe how to interface with an asset through a suitable submodel (e.g., *Asset Interfaces Description* [15]), the actual connection between the AAS and the physical asset remains unspecified and lacks formal guidance. Consequently, this interface is often developed in a customized, case-by-case manner, depending on the specific characteristics and communication protocols of the asset. Furthermore, even when such connections are realized, data exchange is typically unidirectional since AAS Types 1 and 2 do not include the mechanisms required for direct, active, or bidirectional interaction with the physical asset.

In this context, agents are particularly well-suited for interfacing with automation and control devices. Unlike passive data collection mechanisms, agents are intelligent components that not only gather data but also process them to make decisions and provide feedback to the asset. This enables bidirectional communication and real-time adaptability, which are essential for implementing the autonomy and intelligence envisioned in AAS Type 3 solutions. For example, consider an agent associated with a drilling machine. The agent can continuously gather real-time operational data to monitor the drilling process. If it detects an anomaly or suboptimal condition, it can autonomously respond, e.g., by adjusting a control parameter on the machine to correct the issue or by issuing a warning to a supervisory system or human operator.

In order to achieve interoperability and facilitate this integration, it is fundamental to have an easy, transparent, and reusable way of interconnecting agents with different assets. In this context, the IEEE 2660.1 standard [27] presents generic interface practices based on interaction mode and location levels. These interface practices can be used in the context of the defined reference model, particularly to show the different possibilities and guide the selection of the best interface practice to integrate agents with assets, depending on the use case and application requirements.

As illustrated in Figure 6, the IEEE 2660.1 standard defines and compares four integration patterns based on two key factors: the interaction mode between the agent and the asset and the agent's deployment location.

The interaction mode refers to the way the agent interacts with the asset:

- Tightly coupled: This interaction mode is characterized by a direct, permanent, and non-mediated connection between the agent and asset. The participants are connected via direct network communication or shared memory and follow a traditional request–response schema. This schema typically involves the agent sending a request to the asset, which processes the request and sends back a response.
- Loosely coupled: This interaction mode is characterized by an indirect connection between the agent and asset, usually mediated by one or more message brokers, following a publish–subscribe schema. Messages are sent asynchronously, meaning that the agent does not have to wait for an immediate response from the asset.

Regarding the deployment location, two possibilities are identified:

- Hybrid: The agent operates on a separate computational platform from the asset controller.
- On-device: The agent operates on the same computational platform as the asset controller.

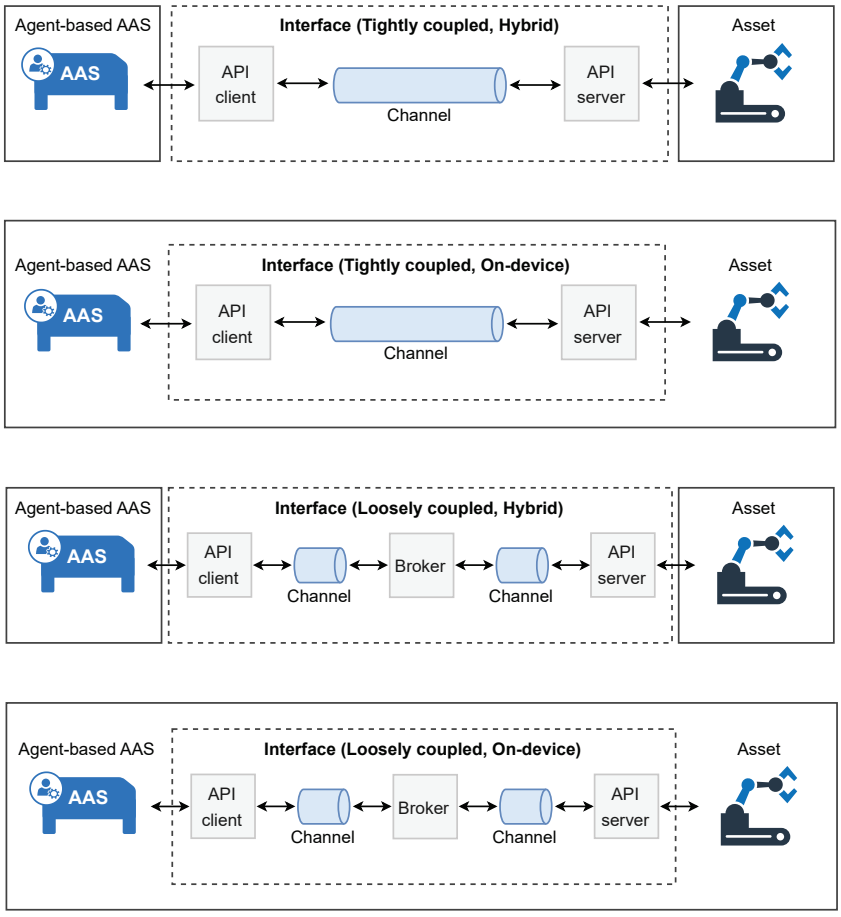


Figure 6. IEEE 2660.1-2020 generic integration patterns (adapted from [27]).

In industrial settings, any proposed solution must be integrated with existing deployed systems. In this context, agent integration can be quite restricted depending on the specifics of the asset, but it can also be flexible, allowing multiple integration patterns and technologies to be applied for the same use case. In such circumstances, it is important to take into account the advantages and disadvantages of each integration pattern for different applications. In this context, the IEEE 2660.1 also defines a recommendation method

that recommends the best interface practices to interconnect agents and physical assets, particularly low-level automation devices, based on user-selected criteria. These criteria include functions (e.g., control, monitoring, or simulation), application domains (e.g., factory automation, building automation, or energy and power systems), technological constraints (e.g., the ability to host agents in the asset controller), as well as the relevant characteristics for the application scenario (e.g., response time, scalability, and reusability). Based on these criteria defined by the user, the method uses a recommendation engine that evaluates existing interface practices stored in a repository, which is continuously updated with interface practices collected from several industrial agent experts who have implemented and used different interface practices. As a result, the recommendation engine produces a ranked list of recommended interface practices, each assigned a score ranging from 0 to 5.

To illustrate the application of the IEEE 2660.1 recommendation method, consider the digitization of an inspection gauge placed in a production line performing inspection tasks to detect deviations in product production. In this context, the application domain is "factory automation", the asset cannot host agents locally, and the functionality provided by the agent is the "control" of the asset, e.g., to adjust its measurement parameters that need to be changed according to the specifications of the product. In such a configuration, "response time" and "reusability" are the most important characteristics to be considered (weighted as response time 80% and reusability 20%). The results provided by the IEEE 2660.1 recommendation method are illustrated in Figure 7.

Results

Recomended interface practice

Id practice	Location	Interaction mode	API client	Channel	Broker	Score
HT-7	Hybrid	Tightly coupled	Java	OPC UA		3.20

Details

Id practice	Location	Interaction mode	API client	Channel	Broker	Score
HT-7	Hybrid	Tightly coupled	Java	OPC UA		3.20
HT-5	Hybrid	Tightly coupled	Apache Milo	OPC-UA		3.20
HT-6	Hybrid	Tightly coupled	Java	Ethernet/IP		2.56
HT-3	Hybrid	Tightly coupled	Java	HTTP		2.56
HT-2	Hybrid	Tightly coupled	Java	Sockets		1.82
HT-4	Hybrid	Tightly coupled	C/C++/SQL	Sockets		1.63
HL-1	Hybrid	Loosely coupled	Apache Paho	MQTT	Eclipse Mosquitto	1.63
HT-1	Hybrid	Tightly coupled	Java	Modbus		1.56
HL-2	Hybrid	Loosely coupled	Java	OPC-UA		1.45
HL-3	Hybrid	Loosely coupled	C/C++/SQL	Sockets	DBMS	1.15
HL-4	Hybrid	Loosely coupled	Apache Paho	MQTT		0.41

Figure 7. Results of the recommended interface practices for the inspection gauge.

The recommended practice ("HT-7") involves implementing a tightly coupled interface utilizing the OPC UA protocol. Due to hardware constraints preventing the hosting of agents on the asset, the suggested strategy relies on a hybrid location setup. This implies that the client operates remotely, potentially leading to response time being influenced by the communication infrastructure. If the asset is not capable of supporting the OPC UA protocol, other alternative interface practices can be considered, e.g., "HT-6" and "HT-3", which consider Ethernet/IP and HTTP protocols, respectively. Based on these considerations, the agent can be designed and implemented in alignment with the characteristics of the recommended interface to meet the desired requirements.

4.4. Communication Based on VDI/VDE 2193 and FIPA Standards

The agent communication capability is another key feature for AAS Type 3 solutions and is essential for enabling seamless, real-time information exchange and collaboration among distributed assets. This capability allows AASs to negotiate, coordinate tasks in a distributed manner, and share knowledge/information, enhancing the flexibility,

reconfigurability, and responsiveness of I4.0 systems. In this context, the adoption of proper standards, such as the I4.0 language, as specified in the VDI/VDE 2193 standards, should be considered to enable interactions between AAS Type 3 solutions. These standards define both the vocabulary and message structures (VDI/VDE 2193-1) [42], as well as the interaction protocols (VDI/VDE 2193-2) [43], which can be used as the basis for agent communication design.

The importance of adopting the I4.0 language as the official communication standard rather than relying on more general agent-specific standards such as FIPA lies in two main reasons: (i) its direct alignment with the definitions and requirements of AAS Type 3, and (ii) although this work employs an agent-based approach to support AAS Type 3 functionalities, it must still be recognized and treated as an AAS Type 3 solution. Therefore, to ensure communication interoperability, it is essential that all solutions conform to common vocabulary, message structure, and interaction protocols, especially considering that other AAS Type 3 implementations may be based on different architectural or technological approaches. Nonetheless, by adopting the I4.0 language as a shared communication standard, these heterogeneous solutions can still interoperate and collaborate effectively, regardless of their internal implementation choices.

Although the I4.0 language is recommended for enabling I4.0-compliant interactions and has a broad scope, the current VDI/VDE 2193-2 standard specifies only the bidding protocol (based on the classic contract net protocol). This limitation restricts solutions that require a wider variety of interaction patterns to support diverse collaboration strategies among AAS Type 3 instances. In this context, several interaction protocols already defined in the FIPA agent communication language (ACL) standards could be adapted to align with the guidelines of the I4.0 language. This transition from FIPA-based interaction protocols to I4.0 language-based protocols is expected to be straightforward, given the strong similarity and compatibility between the two standards. In fact, the VDI/VDE 2193 standards are inspired by the FIPA specifications. For instance, Table 5 maps the parameters of the FIPA ACL message structure to the corresponding elements defined in VDI/VDE 2193-1, demonstrating their equivalence. Furthermore, the communicative acts defined in both standards exhibit only minimal differences, primarily in nomenclature. For a comparison, see [42,44].

In this context, future versions or extensions of the VDI/VDE 2193 standards could formally standardize the mapping between FIPA protocols and the I4.0 language, ensuring their official integration and enabling consistent, interoperable communication. Some examples of candidate FIPA interaction protocols are as follows: request interaction protocol [45], subscribe interaction protocol [46], propose interaction protocol [47], query interaction protocol [48], and request when interaction protocol [49]. While such formal mappings would facilitate development and promote interoperability, it is important to note that the proposed reference model remains implementable using the current FIPA and VDI/VDE 2193 standards independently. However, this may require manual integration efforts, which could limit interoperability and increase development complexity.

Another crucial aspect of communication is vocabulary, which relates to the semantic dimension of information exchange. In the I4.0 context, this information can be contained in individual submodels and should be exchangeable not only within an organization but also across company boundaries. Although these submodels follow a standardized structure defined by the AAS information metamodel, inconsistencies in their semantic interpretation may still arise. To address this, all communication partners must share a common vocabulary. This vocabulary should be based on property-based machine-readable descriptions, i.e., the attributes of an asset in the physical world that denote its characteristics. In this context, the use of standardized dictionaries such as ECLASS [50]

and IEC CDD [51] is recommended, as they enable the description of properties in an unambiguous, machine-readable, and industry-independent manner.

Table 5. Comparison between the message structure of VDI/VDE 2193-1 and FIPA ACL.

VDI/VDE 2193-1				FIPA ACL
Message Area	**Message Element**	**Description**	**Use**	**Equivalent msg. Parameter**
`interaction-Elements`	`-`	Content of a message (submodels or another data elements)	O	`content`
	`type`	Purpose, i.e., intention of a message	M	`performative`
	`sender`	Sender of a message (AAS ID)	M	`sender`
	`receiver`	Receiver of a message (AAS ID)	O	`receiver`
	`conversationId`	ID of a conversation	O	`conversation-id`
`frame`	`messageId`	ID of a message	M	`-`
	`replyTo`	Expression referencing the original message	O	`reply-with`
	`replyBy`	Time by which the response must be received	O	`reply-by`
	`semanticProtocol`	Identification of the used semantic protocol	M	`protocol`
	`role`	Role of the sender of the message	O	`-`

M: mandatory; O: optional.

Finally, to illustrate how agents enable the communication capabilities of an AAS Type 3 solution, Figure 8 presents an example in which agents communicate using the FIPA request interaction protocol, adapted to comply with the I4.0 language guidelines using proper message structures and a shared vocabulary.

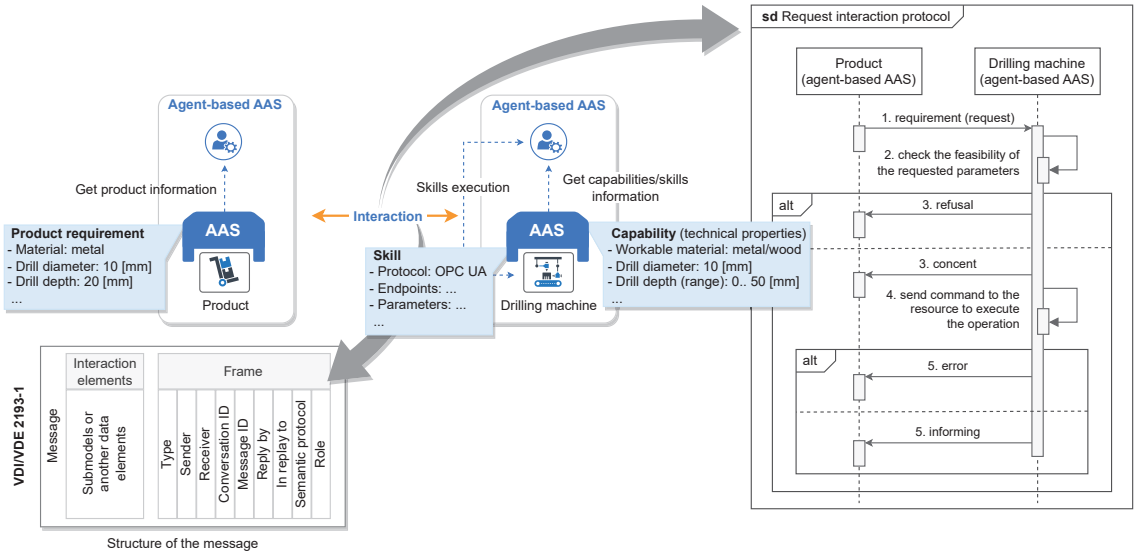


Figure 8. Agent communication capability to enable AAS Type 3 collaboration.

In this example, an agent-based AAS representing a product requests a drilling operation from another agent-based AAS representing a drilling machine. The drilling machine agent then evaluates the feasibility of the requested operation parameters and responds with its capability to perform the task. If the operation is feasible, the agent executes the drilling process and reports the results back to the product agent. Throughout this interaction, all necessary information, such as product requirements (e.g., drilling depth, diameter, and material), interfacing details (e.g., communication protocol, endpoints, and parameters) for executing the drilling operation, and other relevant production-related information, are stored within the respective submodels of each AAS. In this context, the content of the

exchanged messages (see Table 5) is derived from the information contained within the respective submodels (e.g., requirements), ensuring consistency and accuracy in the communication between agents. The messages are structured according to VDI/VDE 2193-1 message structure and can be exchanged in JSON format over appropriate communication protocol technologies.

4.5. Strategies for Developing AI-Based Behaviors

Decision-making is another key element identified in the definitions of AAS Type 3, with intelligence being a necessary requirement to support this process. In this context, agents are inherently intelligent entities capable of perceiving their environment, reasoning, and making autonomous decisions to achieve specific goals. However, implementing this process often requires AI-based strategies, ranging from rule-based mechanisms to more advanced AI techniques, e.g., based on Machine Learning (ML) or Deep Learning (DL).

The implementation of rule-based mechanisms can be simple and direct, typically using "if–then" statements that define clear decision-making logic. However, for more complex, dynamic environments, integrating advanced AI models based on ML/DL into agents significantly enhances their intelligence capabilities. In this sense, this section provides an approach to implementing and deploying such AI models into an agent-based AAS, particularly using standardized, machine-readable interchange formats, such as Predictive Model Markup Language (PMML), Portable Format for Analytics (PFA), and Open Neural Network Exchange (ONNX), which enables the description of AI models in a structured and interoperable manner.

It is important to note that the use of PMML, PFA, ONNX, or other standardized formats is highly recommended but not mandatory. These formats may bring several benefits, such as representing AI models in a standardized manner, enhancing interoperability, facilitating seamless integration across diverse systems, and reducing the time required to make a model production-ready. However, alternative approaches to representing and exchanging AI models may also be adopted, depending on specific requirements or constraints, such as unique application needs or legacy system compatibility.

As illustrated in Figure 9, the agent-based AAS should act as a model consumer, where AI models developed in a training/modeling environment are deployed to the deployment/operational environment for execution. The (re)design phase encompasses two key steps: (i) the *model producer* (e.g., experts, systems, or tools) is responsible for developing, training, and validating AI models, and (ii) the *modeling submodel* handles the conversion and exportation of these models into standardized, machine-readable interchange formats, such as PMML, PFA or ONNX, for inclusion in appropriate submodels (e.g., *AI Dataset* [52], *AI Model Nameplate* [53], and *AI Deployment* [54], which describe the AI lifecycle). On the other hand, the operational phase involves deploying the AI models. During this phase, the *model consumer*, represented by the agent-based AAS, retrieves the AI models described in PMML, PFA, ONNX, or another standardized format from the designated submodels and executes them seamlessly.

This decoupling, combined with the use of standardized, machine-readable interchange formats for AI models, facilitates AI development while enabling modularity, interoperability, and seamless sharing of models across various applications and systems. These standardized formats ensure that AI models, developed in specialized environments using appropriate programming languages and tools, can be deployed in environments that might be different from where the models were initially trained, facilitating their integration into diverse operational settings. Moreover, integrating these standardized AI models into AAS submodels offers an I4.0-compliant approach to describing and representing AI within the I4.0 context.

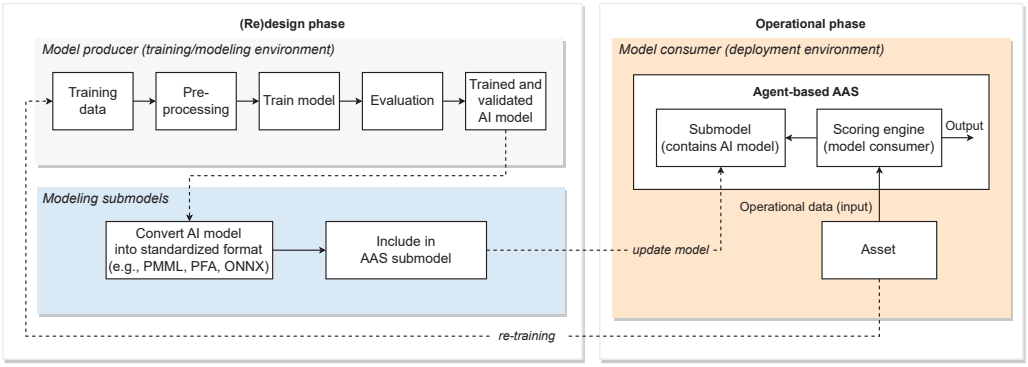


Figure 9. Deployment of AI models into the agent-based AAS.

Based on that, once an AI model is included in a specific submodel, the agent-based AAS needs to retrieve this model from the proper submodel and consume it using an appropriate scoring engine approach. A scoring engine is a software component that processes input data through an AI model to produce, e.g., predictions or inferences. It is able to interpret the specific structure of an AI model and applies the model to input data in real-time. Furthermore, as shown in Figure 9, as new data is generated, the AI model can be retrained and updated within the submodel, ensuring continuous improvement and adaptation to evolving operational conditions.

As an illustrative example, consider the classic Iris dataset translated into a manufacturing context, such as the monitoring of an asset. As shown in Figure 10, the asset dataset consists of various parameters collected from the asset, such as temperature, pressure, vibration, and humidity, recorded over a period of time. These parameters are typically labeled to indicate whether the asset is operating in a normal or faulty state. In a training/modeling environment, the dataset is used to train an AI model based on the random forest algorithm (e.g., utilizing tools or programming languages such as Python, R, or others). Once the model is trained and validated, it is converted into PMML format and included in the *AI Deployment* submodel. Additionally, information about the training process, dataset characteristics, requirements, and other relevant details are stored within the *AI Dataset* and *AI Model Nameplate* submodels.

In the operational phase, the agent-based AAS, which may be developed in a Java environment, retrieves the model from the *AI Deployment* submodel and consumes it using a scoring engine (e.g., through Java PMML API). By using this model, the agent-based AAS can predict asset states based on input data and trigger alerts when potential faults are detected. These alerts can inform supervisory systems, human operators, and other agents, enabling them to reorganize workflows and reallocate tasks on the fly. It is important to emphasize that the focus here is not on the AI algorithm itself (e.g., its performance or accuracy) but rather on demonstrating a standardized approach that enables agent-based AAS to consume diverse AI models tailored to specific applications in compliance with I4.0, aiming to implement intelligence in AAS Type 3 solutions.

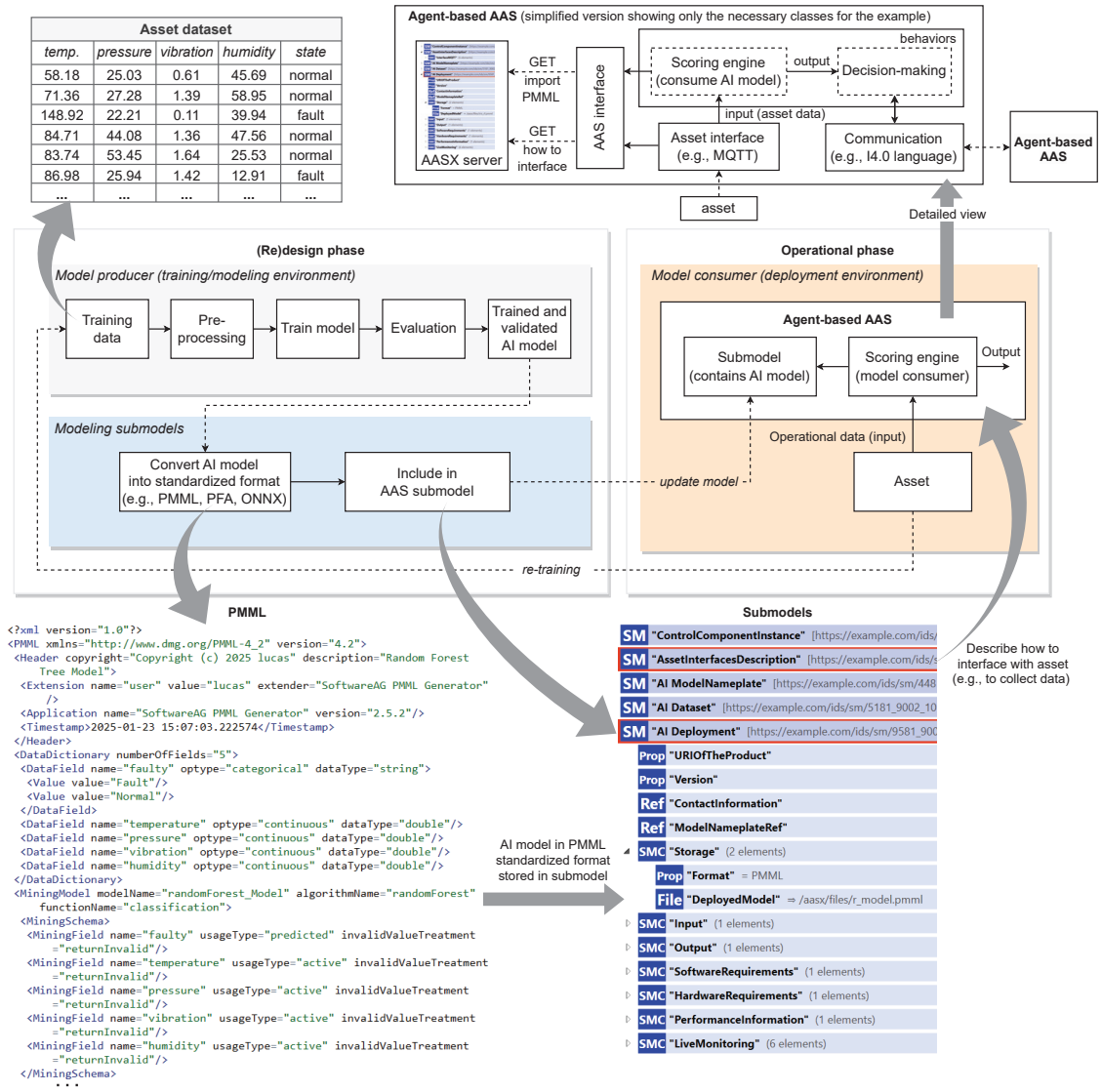


Figure 10. Example of AI model deployment for asset monitoring in agent-based AAS.

5. Conclusions and Future Work

AAS Type 3 plays a pivotal role in realizing the full vision of I4.0, characterized by fully interoperable, collaborative, intelligent, resilient, and adaptive industrial ecosystems. Despite its significance, AAS Type 3 remains in an early stage of development and still lacks comprehensive specifications and practical implementation guidelines. In this context, MAS have been investigated as a means to enhance traditional AAS solutions toward the realization of AAS Type 3, particularly by embedding autonomous, intelligent, and collaborative behaviors.

This paper explored the role of MAS in realizing AAS Type 3 through a comprehensive analysis of existing agent-based AAS approaches in the literature. In summary, MAS enable traditional AAS solutions to be extended with active and intelligent behaviors that support autonomous decision-making and collaboration in production processes. The reviewed literature highlights solutions across various application domains, namely smart manufacturing, dynamic production replanning, shared production, control, scheduling, human collaboration, diagnosis, and fault detection. However, none of the identified approaches have yet been deployed or tested in real-world scenarios, underscoring the early stage and limited practical validation of agent-based AAS solutions. On the other hand, the promising results from simulation and prototype experiments indicate significant

potential for these approaches to improve interoperability, flexibility, scalability, responsiveness, and intelligence in industrial systems. Furthermore, most studies are focused on addressing a particular use case and do not provide clear guidance for generalizing the approach or developing comprehensive, reusable solutions. Finally, given that this is a relatively new and evolving field, further research and development are essential to consolidate its theoretical and conceptual foundations before transitioning from the state of the art to practical implementation in industrial environments.

Based on the findings, a reference model was proposed that captures identified recurring patterns and best practices in a unified conceptual model, offering a standardized and modular foundation for understanding and guiding the development of AAS Type 3 solutions. The proposed model extends the existing AAS structure [10], facilitating the seamless transition from established Type 1 and Type 2 implementations toward more autonomous and intelligent Type 3 solutions. Moreover, as a reference model, it serves as a template that provides a high-level structure from which more detailed and concrete architectures and specifications can be derived and tailored to specific industrial applications and requirements.

Although the current version of the proposed reference model is conceptual in nature and has not yet been empirically validated, its structure and design guidelines are conceptually grounded in a formal analysis of existing agent-based AAS solutions reported in the literature. All these solutions have been validated through simulations, prototypes, or conceptual evaluations, as discussed in Section 3.

Future work will focus on developing a set of specifications for designing and implementing AAS Type 3 using an agent-based approach. These specifications will build upon the proposed reference model, providing more detailed technical guidance for both architectural design and practical implementation across various industrial scenarios.

Moreover, empirical validation will be pursued through complementary strategies. First, a prototype will be implemented based on a representative use case, such as negotiation among AASs in a production line, to demonstrate feasibility. Second, the model will be applied in a realistic case study, potentially in collaboration with an industrial partner, to assess its performance, scalability, and adaptability in practice. These efforts will serve to confirm the model's applicability and support its iterative refinement based on practical insights.

Author Contributions: Conceptualization, L.S.; methodology, L.S.; software, L.S.; investigation, L.S.; writing—original draft preparation, L.S.; writing—review and editing, L.S., F.D.l.P. and P.L.; supervision, F.D.l.P. and P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by national funds: UID/05757—Research Centre in Digitalization and Intelligent Robotics (CeDRI), and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020). Lucas Sakurada thanks the FCT for the PhD Grant 2020.09234.BD (DOI: 10.54499/2020.09234.BD).

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Plattform Industrie 4.0. *Structure of the Administration Shell—Continuation of the Development of the Reference Model for the Industrie 4.0 Component*; Plattform Industrie 4.0: Berlin, Germany, 2016.
2. *DIN SPEC 91345*; Reference Architecture Model Industrie 4.0 (RAMI4.0). DIN: Berlin, Germany, 2016. (In English)
3. Plattform Industrie 4.0. *Verwaltungsschale in der Praxis*; Plattform Industrie 4.0: Berlin, Germany, 2020.
4. Wooldridge, M. *An Introduction to MultiAgent Systems: Second Edition*; John Wiley & Sons: Hoboken, NJ, USA, 2009.

5. Karnouskos, S.; Ribeiro, L.; Leitão, P.; Lüder, A.; Vogel-Heuser, B. Key Directions for Industrial Agent Based Cyber-Physical Production Systems. In Proceedings of the 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS), Taipei, Taiwan, 6–9 May 2019; pp. 17–22.
6. Vogel-Heuser, B.; Seitz, M.; Salazar, L.A.C.; Gehlhoff, F.; Dogan, A.; Fay, A. Multi-agent systems to enable Industry 4.0. *at-Automatisierungstechnik* **2020**, *68*, 445–458. [CrossRef]
7. Sakurada, L.; Leitão, P. Multi-Agent Systems to Implement Industry 4.0 Components. In Proceedings of the 2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS), Tampere, Finland, 10–12 June 2020; Volume 1, pp. 21–26.
8. Leitão, P.; Karnouskos, S.; Ribeiro, L.; Lee, J.; Strasser, T.; Colombo, A.W. Smart Agents in Industrial Cyber–Physical Systems. *Proc. IEEE* **2016**, *104*, 1086–1101. [CrossRef]
9. Karnouskos, S.; Leitão, P.; Ribeiro, L.; Colombo, A.W. Industrial Agents as a Key Enabler for Realizing Industrial Cyber-Physical Systems: Multiagent Systems Entering Industry 4.0. *IEEE Ind. Electron. Mag.* **2020**, *14*, 18–32. [CrossRef]
10. Industrial Digital Twin Association. *Specification of the Asset Administration Shell—Part 1: Metamodel*; Version 3.0.1; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2024.
11. Nakagawa, E.Y.; Antonino, P.O.; Schnicke, F.; Capilla, R.; Kuhn, T.; Liggesmeyer, P. Industry 4.0 reference architectures: State of the art and future trends. *Comput. Ind. Eng.* **2021**, *156*, 107241. [CrossRef]
12. Industry IoT Consortium. *The Industrial Internet Reference Architecture*; Industry IoT Consortium: Boston, MA, USA, 2022.
13. Industry Internet Consortium. Plattform Industrie 4.0. *Architecture Alignment and Interoperability*; Industry Internet Consortium: Boston, MA, USA; Plattform Industrie 4.0: Berlin, Germany, 2017.
14. Industrial Digital Twin Association. *Specification of the Asset Administration Shell—Part 2: Application Programming Interfaces*; Version 3.0.3; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2024.
15. Industrial Digital Twin Association. *IDTA 02017-1-0 Asset Interfaces Description*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2024.
16. Wagner, C.; Grothoff, J.; Epple, U.; Drath, R.; Malakuti, S.; Grüner, S.; Hoffmeister, M.; Zimermann, P. The role of the Industry 4.0 asset administration shell and the digital twin during the life cycle of a plant. In Proceedings of the 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 12–15 September 2017.
17. Abdel-Aty, T.A.; Negri, E.; Galparoli, S. Asset Administration Shell in Manufacturing: Applications and Relationship with Digital Twin. *IFAC-PapersOnLine* **2022**, *55*, 2533–2538. [CrossRef]
18. Zhang, J.; Ellwein, C.; Heithoff, M.; Michael, J.; Wortmann, A. Digital twin and the asset administration shell. *Softw. Syst. Model.* **2025**, *24*, 771–793. [CrossRef]
19. Plattform Industrie 4.0. *Functional View of the Asset Administration Shell in an Industrie 4.0 System Environment*; Plattform Industrie 4.0: Berlin, Germany, 2021.
20. Plattform Industrie 4.0. *What Is the Asset Administration Shell from a Technical Perspective?* Plattform Industrie 4.0: Berlin, Germany, 2021.
21. Industrial Digital Twin Association. *Specification of the Asset Administration Shell—Part 4: Security*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2025.
22. Ye, X.; Song, W.S.; Hong, S.H.; Kim, Y.C.; Yoo, N.H. Toward Data Interoperability of Enterprise and Control Applications via the Industry 4.0 Asset Administration Shell. *IEEE Access* **2022**, *10*, 35795–35803. [CrossRef]
23. Stolze, M.; Belyaev, A.; Kosel, C.; Diedrich, C.; Barnard, A. Realizing Automated Production Planning via Proactive AAS and Business Process Models. In Proceedings of the 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA), Padova, Italy, 10–13 September 2024; pp. 1–8.
24. López, A.; Casquero, O.; Estévez, E.; Armentia, A.; Orive, D.; Marcos, M. An industrial agent-based customizable platform for I4.0 manufacturing systems. *Comput. Ind.* **2023**, *146*, 103859. [CrossRef]
25. Grunau, S.; Redeker, M.; Göllner, D.; Wisniewski, L. The Implementation of Proactive Asset Administration Shells: Evaluation of Possibilities and Realization in an Order Driven Production. In Proceedings of the Kommunikation und Bildverarbeitung in der Automation, Lemgo, Germany, 28–29 October 2022; Jasperneite, J., Lohweg, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2022; pp. 131–144.
26. Vogel-Heuser, B.; Ocker, F.; Scheuer, T. An approach for leveraging Digital Twins in agent-based production systems. *at-Automatisierungstechnik* **2021**, *69*, 1026–1039. [CrossRef]
27. *IEEE Std 2660.1-2020*; IEEE Recommended Practice for Industrial Agents: Integration of Software Agents and Low-Level Automation Functions. IEEE: New York, NY, USA 2021; pp. 1–43.
28. Monostori, L.; Váncza, J.; Kumara, S. Agent-Based Systems for Manufacturing. *CIRP Ann.* **2006**, *55*, 697–720. [CrossRef]
29. López, A.; Casquero, O.; Marcos, M. Design patterns for the implementation of Industrial Agent-based AASs. In Proceedings of the 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Victoria, BC, Canada, 10–12 May 2021; pp. 213–218.

30. López, A.; Casquero, O.; Estévez, E.; Leitão, P.; Marcos, M. Towards the generic integration of agent-based AASs and Physical Assets: A four-layered architecture approach. In Proceedings of the 2021 IEEE 19th International Conference on Industrial Informatics (INDIN), Palma de Mallorca, Spain, 21–23 July 2021; pp. 1–6.
31. Sakurada, L.; Leitao, P.; De La Prieta, F. Engineering a Multi-agent Systems Approach for Realizing Collaborative Asset Administration Shells. In Proceedings of the 2022 IEEE International Conference on Industrial Technology (ICIT), Shanghai, China, 22–25 August 2022; pp. 1–6.
32. Jungbluth, S.; Hermann, J.; Motsch, W.; Pourjafarian, M.; Sidorenko, A.; Volkmann, M.; Zoltner, K.; Plociennik, C.; Ruskowski, M. Dynamic Replanning using Multi-Agent Systems and Asset Administration Shells. In Proceedings of the 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA), Stuttgart, Germany, 6–9 September 2022; pp. 1–8.
33. Sidorenko, A.; Motsch, W.; van Bekkum, M.; Nikolakis, N.; Alexopoulos, K.; Wagner, A. The MAS4AI framework for human-centered agile and smart manufacturing. *Front. Artif. Intell.* **2023**, *6*, 1241522. [CrossRef] [PubMed]
34. Xia, Y.; Shenoy, M.; Jazdi, N.; Weyrich, M. Towards autonomous system: Flexible modular production system enhanced with large language model agents. In Proceedings of the 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA), Sinaia, Romania, 12–15 September 2023; pp. 1–8.
35. Sakurada, L.; De La Prieta, F.; Leitao, P. A Methodology for Integrating Asset Administration Shells and Multi-agent Systems. In Proceedings of the 2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE), Helsinki, Finland, 19–21 June 2023; pp. 1–6.
36. Bernhard, A.T.; Jungbluth, S.; Karnoub, A.; Sidorenko, A.; Motsch, W.; Wagner, A.; Ruskowski, M. *I4.0 Holonic Multi-Agent Testbed Enabling Shared Production*; Springer Nature: Cham, Switzerland, 2024; pp. 231–250.
37. Siatras, V.; Bakopoulos, E.; Mavrothalassitis, P.; Nikolakis, N.; Alexopoulos, K. Production Scheduling Based on a Multi-Agent System and Digital Twin: A Bicycle Industry Case. *Information* **2024**, *15*, 337. [CrossRef]
38. Rübel, P.; Motsch, W.; Bernhard, A.; Jungbluth, S.; Ruskowski, M. Agent-Based Communication for Fault Diagnosis in Skill-Based Production Environments Using Messages Based on I4.0 Language and Asset Administration Shells. In Proceedings of the Advances in Artificial Intelligence in Manufacturing II, Athens, Greece, 16 October 2025; pp. 157–170.
39. Industrial Digital Twin Association. *IDTA 02016-1-0—Control Component Instance*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2023.
40. Industrial Digital Twin Association. *IDTA 02008-1-1 Time Series Data*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2023.
41. Köcher, A.; Belyaev, A.; Hermann, J.; Bock, J.; Meixner, K.; Volkmann, M.; Winter, M.; Zimmermann, P.; Grimm, S.; Diedrich, C. A reference model for common understanding of capabilities and skills in manufacturing. *at-Automatisierungstechnik* **2023**, *71*, 94–104. [CrossRef]
42. *VDI/VDE 2193*; Part 1—Language for I4.0 Components - Structure of Messages. VDI/VDE: Düsseldorf, Germany, 2020.
43. *VDI/VDE 2193*; Part 2—Language for I4.0 components - Interaction protocol for bidding procedures. VDI/VDE: Düsseldorf, Germany, 2020.
44. Foundation for Intelligent Physical Agents. *FIPA Communicative Act Library Specification*; Foundation for Intelligent Physical Agents: Geneva, Switzerland, 2002.
45. Foundation for Intelligent Physical Agents. *FIPA Request Interaction Protocol Specification*; Foundation for Intelligent Physical Agents: Geneva, Switzerland, 2002.
46. Foundation for Intelligent Physical Agents. *FIPA Subscribe Interaction Protocol Specification*; Foundation for Intelligent Physical Agents: Geneva, Switzerland, 2002.
47. Foundation for Intelligent Physical Agents. *FIPA Propose Interaction Protocol Specification*; Foundation for Intelligent Physical Agents: Geneva, Switzerland, 2002.
48. Foundation for Intelligent Physical Agents. *FIPA Query Interaction Protocol Specification*; Foundation for Intelligent Physical Agents: Geneva, Switzerland, 2002.
49. Foundation for Intelligent Physical Agents. *FIPA Request When Interaction Protocol Specification*; Foundation for Intelligent Physical Agents: Geneva, Switzerland, 2002.
50. Plattform Industrie 4.0; ECLASS. *Modelling the Semantics of Data of an Asset Administration Shell with Elements of ECLASS*; Plattform Industrie 4.0: Berlin, Germany; ECLASS: Cologne, Germany, 2021.
51. *IEC 61360-4*; IEC/SC 3D—Common Data Dictionary (CDD—V2.0018.0001). International Electrotechnical Commission: London, UK, 2005.
52. Industrial Digital Twin Association. *IDTA 02058-1-0 Artificial Intelligence Dataset*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2025.

53. Industrial Digital Twin Association. *IDTA 02060-1-0 Artificial Intelligence Model Nameplate*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2025.
54. Industrial Digital Twin Association. *IDTA 02059-1-0 Artificial Intelligence Deployment*; Industrial Digital Twin Association: Frankfurt am Main, Germany, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

future internet

MDPI

Article

Implementation and Performance Analysis of an Industrial Robot's Vision System Based on Cloud Vision Services

Ioana-Livia Stefan [1,†], Andrei Mateescu [1,*,†], Ionut Lentoiu [2], Silviu Raileanu [2], Florin Daniel Anton [2], Dragos Constantin Popescu [1] and Ioan Stefan Sacala [1,*]

[1] Automatic Control and Systems Engineering Department, Faculty of Automatic Control and Computers, National University of Science and Technology Politehnica of Bucharest, Splaiul Independentei, No. 313, 060042 Bucharest, Romania
[2] Automatics and Industrial Informatics Department, Faculty of Automatic Control and Computers, National University of Science and Technology Politehnica of Bucharest, Splaiul Independentei, No. 313, 060042 Bucharest, Romani
* Correspondence: andrei.mateescu0612@upb.ro (A.M.); ioan.sacala@upb.ro (I.S.S.)
† These authors contributed equally to this work.

Abstract

With its fast advancements, cloud computing opens many opportunities for research in various applications from the robotics field. In our paper, we further explore the prospect of integrating Cloud AI object recognition services into an industrial robotics sorting task. Starting from our previously implemented solution on a digital twin, we are now putting our proposed architecture to the test in the real world, on an industrial robot, where factors such as illumination, shadows, different colors, and textures of the materials influence the performance of the vision system. We compare the results of our suggested method with those from an industrial machine vision software, indicating promising performance and opening additional application perspectives in the robotics field, simultaneously with the continuous improvement of Cloud and AI technology.

Keywords: ABB RobotStudio; industrial robots; cloud computing; object recognition; Microsoft Azure Custom Vision; computer vision; digital twin

1. Introduction

The desire to improve the performance of manufacturing processes led to the emergence of many novel approaches to keep up with the rapidly evolving world of robotics, one of them being Cloud Robotics [1]. Moving computationally intensive tasks to the cloud enhances a system's performance and computational capabilities, mitigating the known limitations of robotic devices (some notable examples being limited storage and on-board computation, as well as battery and storage capacities) while offering a way of sharing services [2].

In our previous article [3],we utilized a digital twin [4] from ABB RobotStudio in order to test the performance of our proposed solution in a simulated environment before implementing it on the real robot. In this article, we take the proposed architecture from our previous work and compare the results we obtained there in Robot Studio with the results we have obtained now, in the real world, using the ABB IRB 140 industrial robot. We compare the performance obtained with the proposed method, based on Custom Vision, with a commercial machine vision software used in industrial applications. Custom Vision is part of Microsoft Azure's AI services and it allows users to create, train, and

evaluate their own models for image classification or object detection and recognition. We particularly suggest this method as a fallback option for cases where specialized machine vision software is unable to detect objects, such as when they are dirty or overlapping.

2. Literature Review

Various types of AI techniques are being used by the scientific community for classification and detection tasks [5–7]. In the study of [8], cloud technology was used in a face recognition application based on Deep Learning algorithms in order to improve speed and overcome robot onboard limitations. The authors encrypted the images used for training and testing, and tested several deep learning algorithms for training in order to measure and compare the security, time complexity, and accuracy of recognition for algorithms in the cloud and on robots, obtaining the overall best security and recognition accuracy with the genetic algorithm. Their results show that by using the cloud environment, compared to performing the same computation for the same algorithm directly on the robot, the recognition time decreased from 0.559 s to 0.014 s, and the performance of the algorithm was maintained. Therefore, this paper showcases the cloud's ability to decrease the computation time of resource-intensive robotics tasks.

The study [9] presents an autonomous pick-and-sort operation in industrial cloud robotics based on object recognition and motion planning. The object recognition uses a CNN, and to perform the training, cloud computing resources were utilized to meet the high computational resource needs. The CNN method, compared to a CG algorithm, improved the object recognition time by a factor of ten. This application has a similar use case to ours, since they are also executing a pick-and-sort operation using an industrial robot. Their results showed that for a mean inter-object spawning time of 14 s and a mean service time needed for the robot arm to perform the movement of 13 sec, the pick-and-sort operation could reach an overall success rate 71.3%.

CNNs have been used in the fields of artificial vision, object recognition, and natural language processing with exceptional performance. Created based on natural vision systems and neurons, this NN can extract features from data, like images and videos, employing convolutional computations and pooling operations. As one of the main algorithms of deep learning, CNNs are composed of five layers: data input, convolutional computation, activation function, pooling and fully connected layer. In the first layer, the image is preprocessed for speeding up the operations performed in the next layers. The convolutional computation layer is the principal layer in feature extraction, performing several convolution computations to extract various input features. Next, the activation function layer uses numerous activation functions, such as ReLU, ELU, SELU, Sigmoid, Tanh, and many others to map the outputs of the convolutional layer. Then, to reduce the number of features while maintaining the key features, the pooling layer is used. Finally, by modifying parameters and weights, the fully connected layer converts the bidimensional feature map, produced by convolution, into a unidimensional vector, which represents the desired information [10].

To obtain such a CNN, one must train it, which is a time-consuming and a computationally and energy-intensive process that usually takes place using expensive computational resources, like GPUs, which aid this process by parallelization to reduce training time and costs. CNN training presumes a fine-tuning of the layers' internal parameters, known as weights, that determine the layers' outputs and is accomplished in multiple iterations, involving forward passes and backpropagation, loss computations, optimizer updates, and data transmission between the host memory and the GPU. When performing the training, images are processed in batches, the quantity of images processed in a single iteration is called mini-batch size, and precisely one processing of all training images forms a training

epoch. The training of NNs is not an easy feat and researchers are working continuously to better estimate CNN training time and improve network generalization and recognition abilities [11,12].

Cloud computing plays an important role in future Internet-based solutions, supporting the evolution of technology through innovations such as the development of smart cities [13] or collaborative computing [14]. We see that various cloud-based solutions are being approached by researchers for various robotics applications, at different levels of the architecture, while at the same time, effort is being put into improving resources [15]; workload balancing abilities [16]; the mitigation of security vulnerabilities, like threats and attacks [17], to provide information security when accessing information from the cloud [18]; cloud scalability and latency [19]; and the consistency of connectivity of the cloud.

In particular, the study of [20] focused on improving the load balancing abilities of the cloud. They propose a solution to increase the computation performance and resource usage efficiency of the platform by enhancing the resources management module with the use of a flexible container capacity scheduling algorithm. The algorithm performs capacity elastic scheduling operations based on the usage of CPU, memory, and network of the containers. This paper shows how consistent research efforts channeled towards the load-balancing abilities of the cloud while multiple robots have access to the cloud service on demand, could increase the collaborative capabilities of robots.

Another area which is receiving attention is the consistency of connectivity. For example, in [21], an IoRT and cloud robotics solution is proposed based on a three-level client server architecture. Django was utilized for the development of the AWS Cloud hosted application, demonstrating real-time, bidirectional communication capabilities. The authors evaluated the solutions for multiple physical and simulated robots and obtained good capabilities for long-distance global remote control and monitoring using the Internet, while confirming the robustness of the framework, and obtaining minimal communication delays with superior performance in both local and cloud environments.

The study of [22] discusses advancements in addressing issues such as low latency, the scalability of data storage, computation, and resource management efficiency through the incorporation of AI-based optimization, cloud services and workload management strategies to enhance the Internet-driven applications.

As the capabilities of cloud platforms increase and further improvements are made regarding their resources, workload balancing ability, and the consistency of connectivity between users or devices, we expect to see them integrated in robotics solutions increasingly often.

In the remainder of this paper, we will present the constituent components of our architecture, how the implementation was achieved, the results obtained for two illumination conditions (natural light vs. artificial light) and for both solutions (the proposed solution based on Custom Vision vs. the industrial machine vision software), and our conclusions after analyzing the results.

3. Materials and Methods

Our suggested solution for the sorting task incorporates Microsoft Azure Custom Vision, a cloud service that enables the localization (obtaining the position) and identification (obtaining the type) of items in a picture using a user-created custom model. The position is denoted by a bounding box in which the object is contained, and the type is represented by a tag and a detection confidence returned by the service. The model was trained using 40 images, taken by a camera in good illumination conditions, with plenty of natural light, of 5 objects of different shapes and colors (a black circle with a radius of

16.5 mm; a purple square with sides of 25 mm; a green rhombus with sides of 25 mm; two yellow trapezoids with bases measuring 50 mm and 25 mm, respectively; and a red hexagon with sides of 25 mm), placed in different positions and orientations. The training is performed by assigning bounding boxes and tags for the objects present in the training images, and then an ONNX model is computed by the service. The result of the training process can be seen in Figure 1.

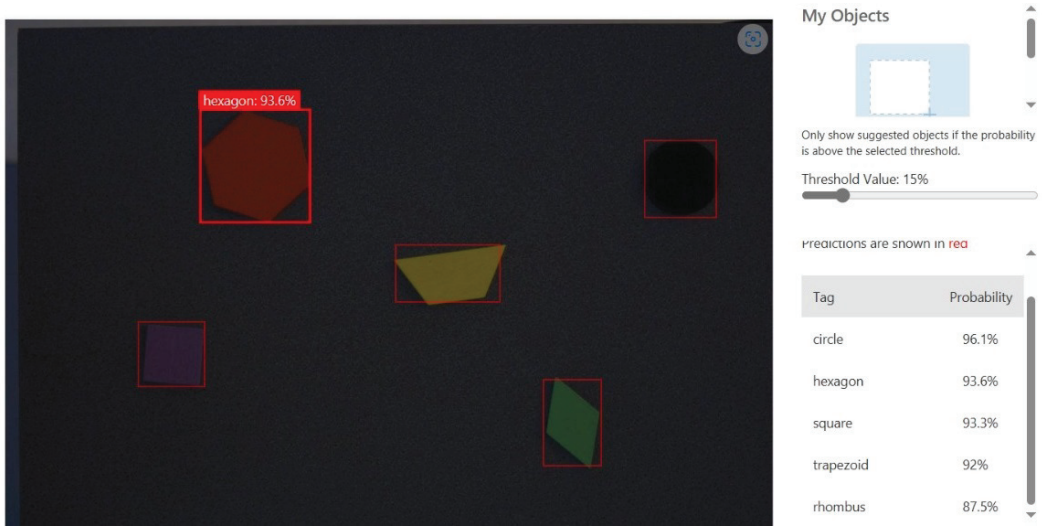


Figure 1. Training result of the model.

The layout of the system is made up of an industrial robot, an IRB 140 from ABB, a vision system, an Ace2 R a2A2590-22gcBAS [23] camera from Basler with a C125-1620-5M f16mm lens [24], the Microsoft Azure Custom Vision cloud service, and a computer that acts as an interface between the previously mentioned components, as shown in Figure 2. The same system is presented in Figure 3a, where the vacuum suction gripper used for manipulating the objects is highlighted.

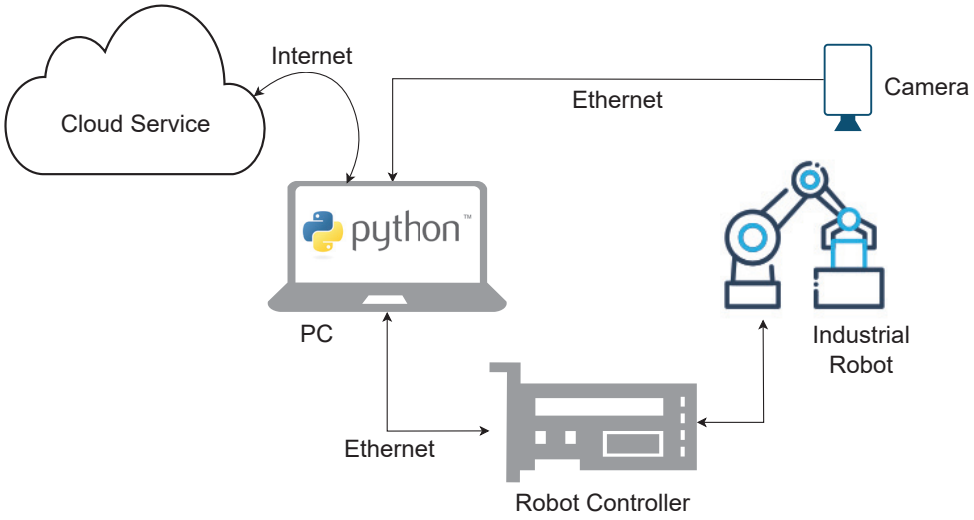
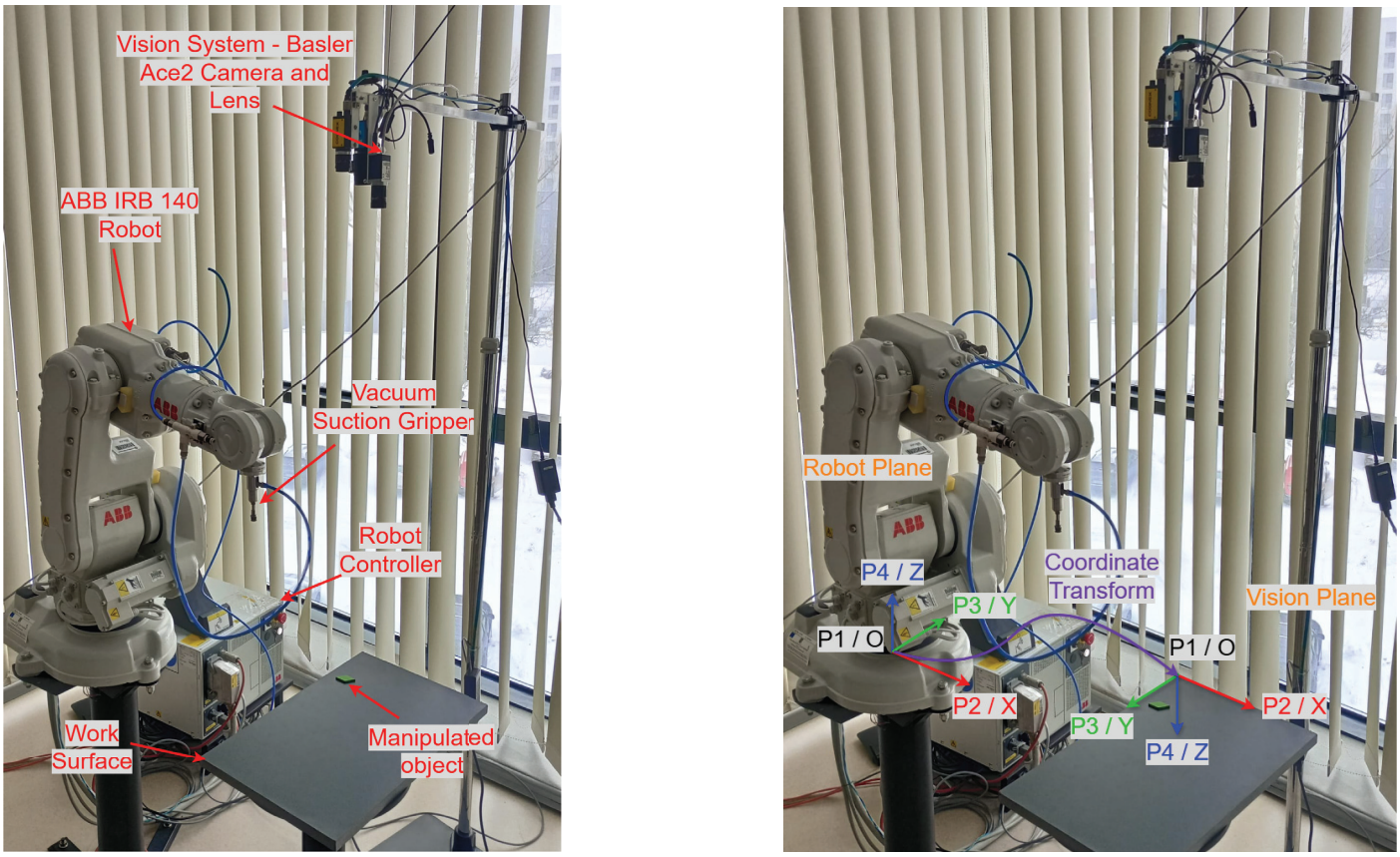


Figure 2. Cloud vision system architecture.

The computer is the centerpiece of the system. It runs a Python script that connects to the camera via a GigE connection, using the Basler pyPylon wrapper and requests images from the camera. The vision system, upon receiving a request, takes a picture and sends it back to the computer, where the same Python script creates a connection to the cloud and sends the pictures to Microsoft Azure Custom Vision over an API, and then awaits a

response. When the cloud service API response is received, it is processed and transmitted to the robot arm–robot controller pair for the sorting process.

Azure AI Custom Vision provides an API answer in JSON format that includes the kind of item identified, the position, indicated by a bounding box, and the detection probability. Following camera calibration, the object location in camera frame, in millimeters, is calculated using the pixel-to-millimeter ratio. Applying a coordinate system transform (Figure 3b), the position of the handling point in the robot frame is obtained and finally transmitted to the robot controller over the Ethernet connection by the Python script. The coordinate system (P1P3P2) for robot–camera calibration is adjusted in the vision system based on the points that are learned with the robot. For this, a special target (arrow) is used. The event diagram for the whole procedure can be seen in Figure 4.



(**a**) (**b**)

Figure 3. (**a**) The physical system and its components. (**b**) The transformation from the robot plane to the vision plane.

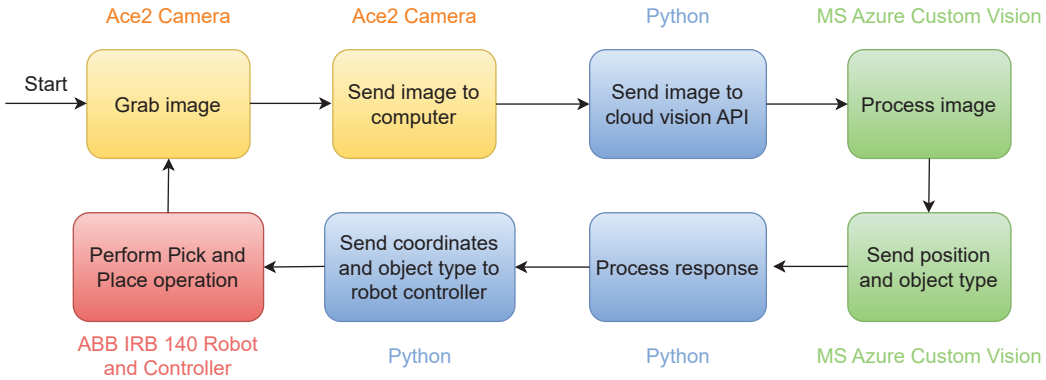


Figure 4. The event diagram.

Experiments were conducted through a material handling task executed by the already introduced industrial robot, which uses guidance information from two sources:

1. The proposed machine vision service;
2. A commercial machine vision software, namely MVTec Merlic.

Both vision systems use the same acquisition solution, namely a fixed network industrial camera (the GIGE type from Basler) that inspects the vision plane. The handling principle, presented in (Figure 3b), uses the concept of workobject which is a reference frame used to define the position and orientation of a workpiece or fixture in space. In our case, the workobject is in the vision plane. The robot–camera or hand–eye calibration (ref) consists of the following:

1. Aligning the axis of the vision plane with the axis of the robot workobject;
2. Overlaying the origin.

In this respect, after applying the camera calibration described by the translation vector t_{cam}, the rotation matrix R_{cam} (Equations (1) and (2)) and the pixel to mm ratio of 7.942, which transform the position from the camera coordinates to the workobject coordinates, as illustrated in Figure 5, the vision system offers information in mm about the location of the recognized handled object in the learned workobject. This information is used to perform the handling of the object.

$$t_{cam} = \begin{bmatrix} 104 \\ 67 \end{bmatrix}, \qquad R_{cam} = \begin{bmatrix} 0.9996845 & -0.0251177 \\ 0.0251177 & 0.9996845 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} Y_{Wobj} \\ X_{Wobj} \end{bmatrix} = R_{cam} * \left(\begin{bmatrix} X_{Cam} \\ Y_{Cam} \end{bmatrix} - t_{cam} \right) \tag{2}$$

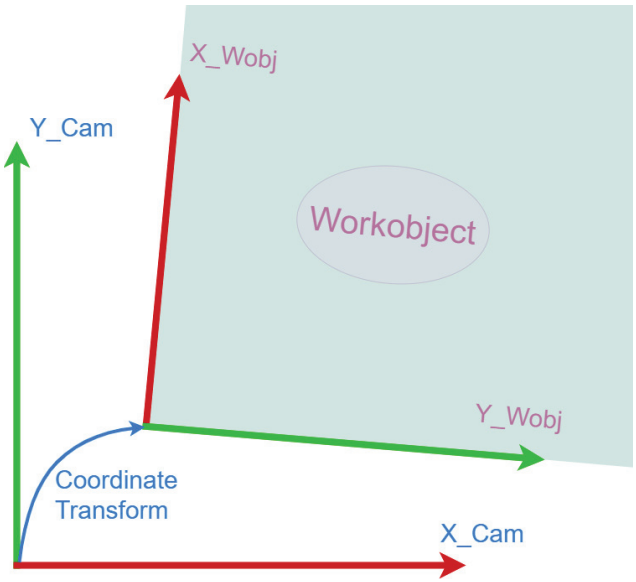


Figure 5. Transforming from camera coordinates to workobject coordinates.

The workobject is defined by the translation vector t and the rotation quaternions q, as seen in Equation (3). The transformation matrix T is computed as Equation (4) and can be used to transform a point from the vision plane to the robot plane as shown in Equation (5), where p_v is the point in the vision plane, and p_r is the point in the robot plane. The inverse operation can be performed as in Equation (6) using the inverse of the transformation matrix T^{-1}, as seen in Equation (7).

$$t = \begin{bmatrix} 511.899 \\ 346.439 \\ 158.981 \end{bmatrix}, \qquad q = \begin{bmatrix} 0.0036446 \\ 0.999819 \\ 0.0160233 \\ -0.00954639 \end{bmatrix} \tag{3}$$

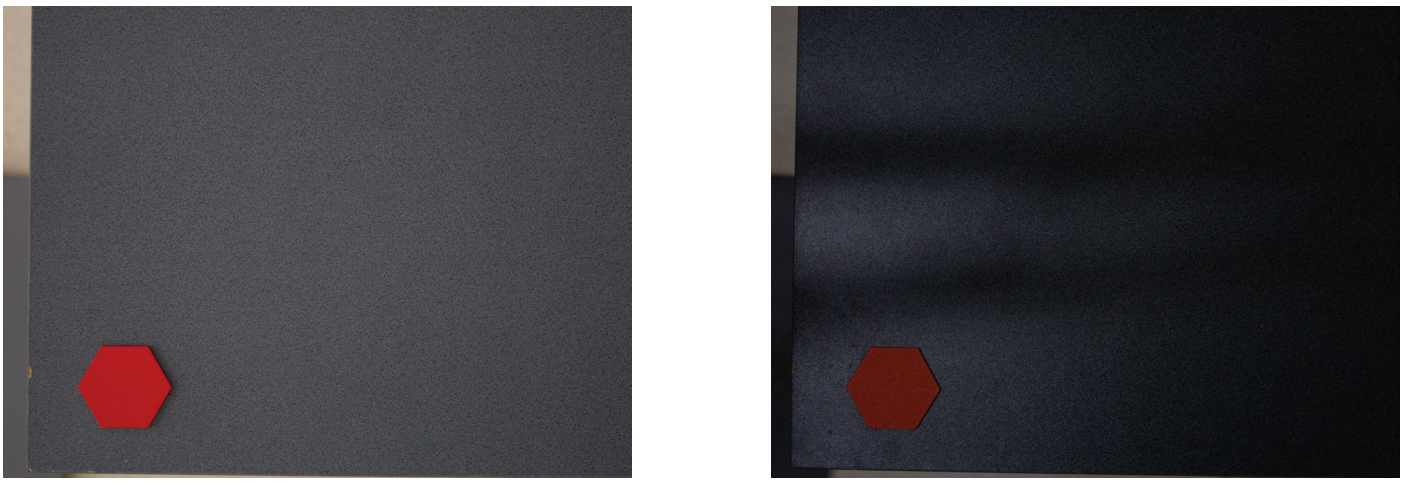
$$T = \begin{bmatrix} 0.99930263 & 0.03211039 & -0.01897253 & 511.899 \\ 0.03197121 & -0.99945994 & -0.00759381 & 346.439 \\ -0.01920612 & 0.00698195 & -0.99979117 & 158.981 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} p_r \\ 1 \end{bmatrix} = T \times \begin{bmatrix} p_v \\ 1 \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} p_v \\ 1 \end{bmatrix} = T^{-1} \times \begin{bmatrix} p_r \\ 1 \end{bmatrix} \tag{6}$$

$$T^{-1} = \begin{bmatrix} 0.9993058 & 0.0319712 & 0.0192061 & 511.566 \\ 0.0321104 & -0.9994599 & 0.0069819 & 328.704 \\ -0.0189725 & -0.0075938 & 0.9997911 & 171.2906 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

The experiments consist of evaluating robot-vision accuracy through both methods, Microsoft Azure Custom Vision and MVTec Merlic, and then comparing the results to see how the cloud service performs against industrial solutions. A robot trajectory, which uses 12 points on the vision plane, was implemented. The robot put the test target (which can be one of the five objects) on the vision plane and the coordinates were stored. The target was left on the vision plane, and its position was computed using the vision system. The image processing time was also measured at the consumer side, from the time a command was issued to the moment the result was received (i.e., from the time the image was sent to the cloud to the moment the results were received). For each object, 12 experiments were performed, under different illuminations, first in similar conditions to the images used for training, i.e., with natural light and then under artificial light. The difference in lighting can be seen in Figure 6 Comparative results are illustrated in the next section.



(**a**) (**b**)

Figure 6. (**a**) Image taken of the hexagonal object under natural light. (**b**) Image taken of the hexagonal object under artificial light.

4. Results

The position of the objects, which are compared in this section, represents the center of the bounding box containing the objects, as seen in Figure 7, and is also the picking point for handling the object.

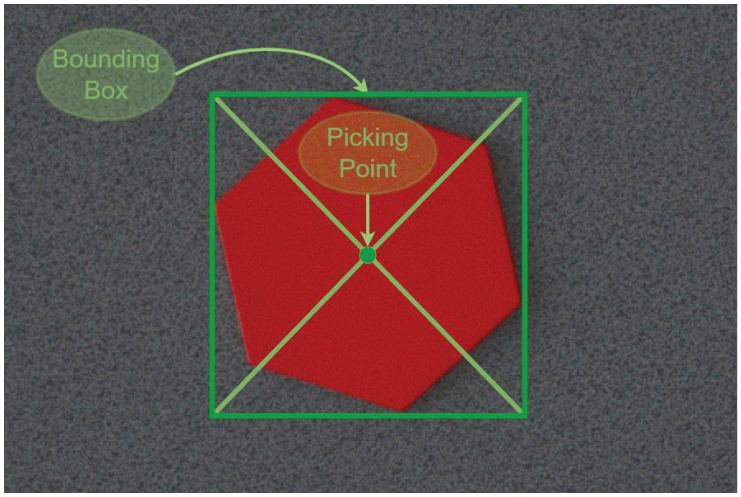


Figure 7. The picking point, calculated as the center of the bounding box.

A screenshot of the bounding boxes and tags for the object that were identified using the industrial solution is presented in Figure 8. The tool for identification is called "Deep Learning AI - Find Objects" and uses an NN computed by the MVTec Deep Learning Tool, trained in an analogous manner to the cloud solution, using the same 40 images of the objects, for which bounding boxes and tags were assigned.

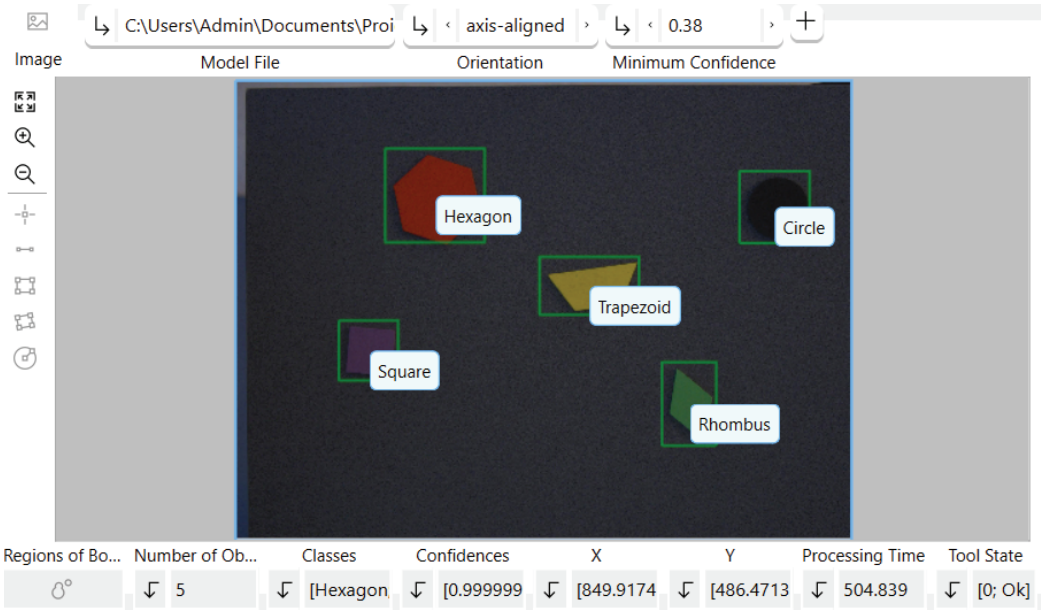


Figure 8. Detected objects, as seen in MVTec Merlic.

In Tables 1 and 2, ten experiments are presented for each system, five experiments conducted under natural light and five under artificial light, with one of each type of object. The detection confidence is displayed as provided by the vision system; the errors represent the position difference in millimeters in the vision plane from the placing point, as used by the industrial robot, and the pick point returned by the vision system. Finally, the time

represents the duration of image processing in seconds or milliseconds, from the moment the image is sent to the vision system to the moment the results are received.

In Tables 3 and 4 the mean values and SD are illustrated, in the same conditions as those previously obtained for each system. For this analysis, twelve experiments were performed for each object, placed in distinct positions and under both types of illumination, totaling 120 experiments.

In Table 3, it can be observed that the lowest detection confidence, when using the cloud service, is obtained for the rhombus, under artificial light, with a mean of 0.63 and an SD of 0.13, while the best detection confidence is gained for the naturally illuminated circle, with a mean of 0.93 and an SD of 0.03. In terms of positioning error, the biggest error is seen when detecting the trapezoid, having a mean error on the X axis of 3.00 mm with an SD of 2.19 mm, and a mean error on the Y axis of 3.13 mm with an SD of 1.35 mm. The smallest errors can be seen in the detection of the circle, with errors of 0.53 mm on the X axis and 0.61 mm on the Y axis, along with SDs of 0.37 mm on the X axis and 0.53 mm on the Y axis.

A similar case can be seen in Table 4, for the industrial application, with the smallest errors for the detection of the circle (0.66 mm mean error on the X axis and 0.4 mm SD and 0.48 mm mean error on the Y axis and 0.29 mm SD), and the biggest error for the trapezoid (3.05 mm mean and 1.69 mm SD on X axis, and 2.89 mm mean and 1.86 SD on Y axis). The confidence in this case varies extraordinarily little, with almost all cases having a confidence of 1, except for the square and trapezoid, which has reduced values when illuminated by artificial light.

Regarding the response time, for both systems, the results are consistent and do not vary much with the object type or ambient light. In the case of the cloud system, the response time varies from a mean of 2.55 s and an SD of 0.14 s in the case of the artificially illuminated hexagon, to a mean of 2.94 s and an SD of 0.39 s for the naturally illuminated circle. The industrial software offers a response in about 500 ms in all cases, with the lowest average value being for the artificially lightened square (mean of 472.6 ms and SD of 27.7 ms) and the highest mean value for the naturally lightened trapezoid (mean of 564.6 ms and SD of 37.3 ms).

Using the data extracted from the same experiments, histograms for confidence (Figures 9 and 10), error on 'X' axis (Figures 11 and 12), error on 'Y' axis (Figures 13 and 14) and time (Figures 15 and 16) have been plotted for both systems. It can be observed that the mean and mode vary heavily for the plotted histograms, which denotes that the distributions are negatively skewed for the confidence, and positively skewed for the errors, due to their limits. The confidence cannot exceed 100%, and the error cannot be less than 0 mm.

Table 1. Table presenting a part of the data obtained when using the cloud solution.

Experiment Number	Ambient Light	Object Type	Detection Confidence	X Error [mm]	Y Error [mm]	Time [s]
1	Natural	Circle	0.95	0.13	1.80	3.50
2	Natural	Square	0.95	1.97	1.41	2.78
3	Natural	Rhombus	0.87	1.87	1.83	2.57
4	Natural	Trapezoid	0.96	0.10	2.90	2.57
5	Natural	Hexagon	0.82	0.17	0.30	2.45
6	Artificial	Circle	0.86	0.66	0.52	2.40
7	Artificial	Square	0.92	2.77	1.89	2.37
8	Artificial	Rhombus	0.78	0.65	0.81	2.87
9	Artificial	Trapezoid	0.79	0.63	3.45	2.45
10	Artificial	Hexagon	0.90	1.37	4.30	2.40

Table 2. Table presenting a part of the data obtained when using the industrial solution.

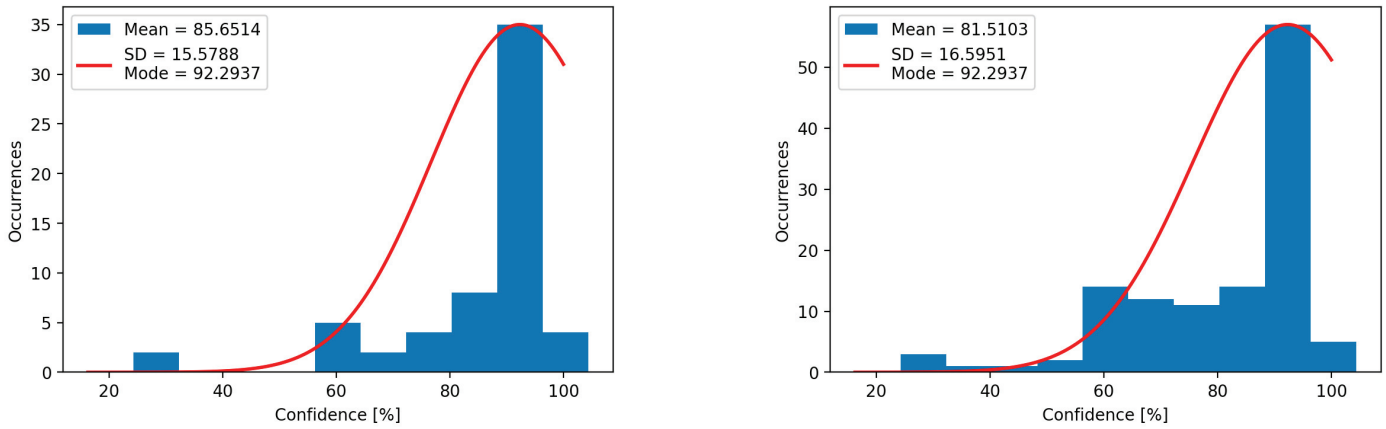
Experiment Number	Ambient Light	Object Type	Detection Confidence	X Error [mm]	Y Error [mm]	Time [ms]
1	Natural	Circle	0.99	2.15	1.04	481.3
2	Natural	Square	0.99	2.73	1.83	563.8
3	Natural	Rhombus	0.99	4.28	0.46	528
4	Natural	Trapezoid	0.99	3.29	1.51	533.6
5	Natural	Hexagon	1	4.07	1.30	520.2
6	Artificial	Circle	1	0.36	0.28	467.5
7	Artificial	Square	0.99	2.88	0.97	463
8	Artificial	Rhombus	0.99	3.23	1.14	474.6
9	Artificial	Trapezoid	0.99	5.60	0.51	459
10	Artificial	Hexagon	0.99	3.53	0.86	532.1

Table 3. The mean and standard deviation for the confidence, error, and response time for all types of objects in different types of light for the cloud product.

Ambient Light	Object Type	Detection Confidence Mean/SD	X Error [mm] Mean/SD	Y Error [mm] Mean/SD	Time [s] Mean/SD
Natural	Circle	0.93/0.03	1.04/0.62	1.71/0.92	2.94/0.39
Natural	Square	0.90/0.08	1.78/1.04	1.49/0.99	2.64/0.19
Natural	Rhombus	0.68/0.17	1.47/1.23	1.46/1.03	2.57/0.20
Natural	Trapezoid	0.85/0.20	2.9/1.79	1.84/1.04	2.57/0.14
Natural	Hexagon	0.89/0.05	2.50/1.85	2.05/1.41	2.64/0.15
Artificial	Circle	0.93/0.03	0.53/0.37	0.61/0.53	2.60/0.25
Artificial	Square	0.92/0.02	1.97/1.16	1.48/0.96	2.68/0.31
Artificial	Rhombus	0.63/0.13	2.06/1.00	1.46/0.91	2.65/0.26
Artificial	Trapezoid	0.66/0.11	3.00/2.19	3.13/1.35	2.75/0.44
Artificial	Hexagon	0.70/0.15	2.57/1.59	1.90/1.10	2.55/0.14

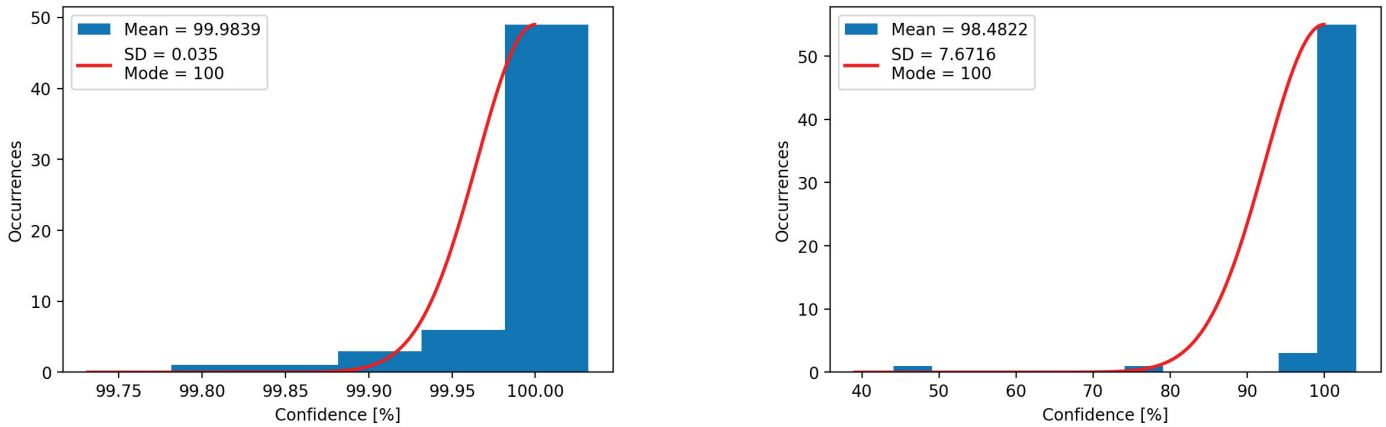
Table 4. The mean and standard deviation for the confidence, error and response time for all types of objects in different types of light for the industrial software.

Ambient Light	Object Type	Detection Confidence Mean/SD	X Error [mm] Mean/SD	Y Error [mm] Mean/SD	Time [ms] Mean/SD
Natural	Circle	0.99/0	1.75/0.77	1.13/1.01	484.5/ 13.9
Natural	Square	0.99/0	2.52/0.85	1.26/0.94	548.6/30.6
Natural	Rhombus	0.99/0	2.6/1.43	1.08/0.75	549.8/20.9
Natural	Trapezoid	0.99/0	3.05/1.69	2.56/1.85	564.6/37.3
Natural	Hexagon	1/0	2.49/1.96	1.84/1.11	553.6/28.3
Artificial	Circle	1/0	0.66/0.40	0.48/0.29	473.5/20
Artificial	Square	0.95/0.15	2.76/1.2	1.29/0.87	472.6/27.7
Artificial	Rhombus	0.99/0.01	2.47/1.07	1.1/1.02	467.4/18.4
Artificial	Trapezoid	0.97/0.06	2.89/2.48	2.89/1.86	473.9/22.4
Artificial	Hexagon	1/0	2.79/1.57	1.46/0.84	482.8/22.3



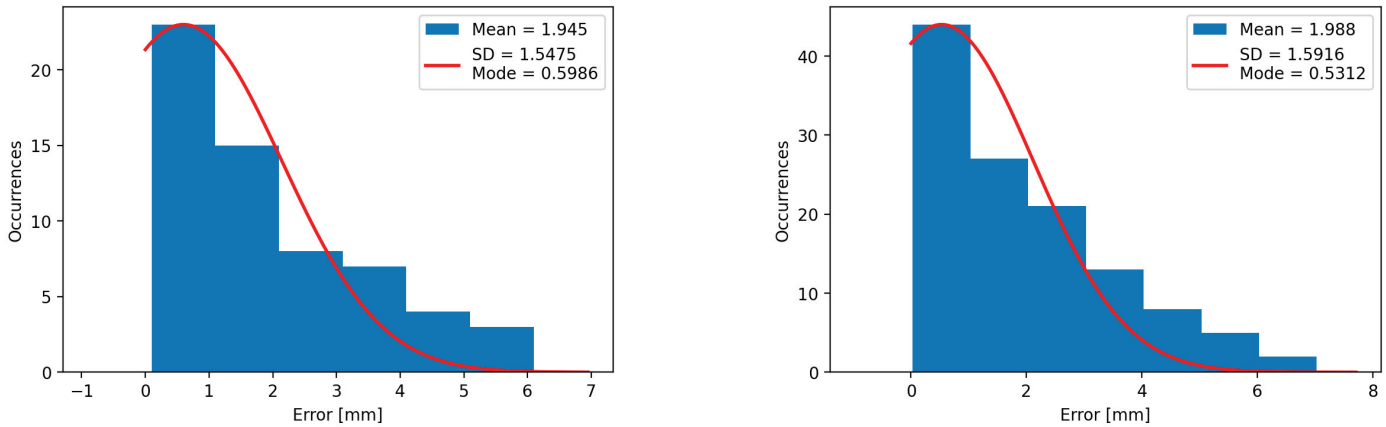
(**a**) (**b**)

Figure 9. (**a**) Confidence histogram for natural light. (**b**) Confidence histogram for artificial light.



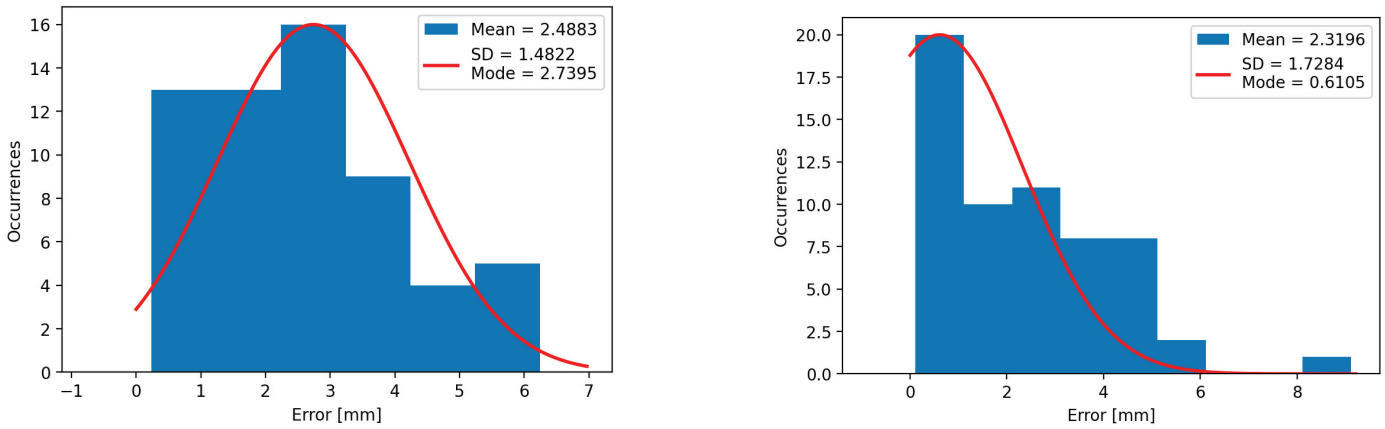
(**a**) (**b**)

Figure 10. (**a**) Confidence histogram for natural light for the industrial software. (**b**) Confidence histogram for artificial light for the industrial software.



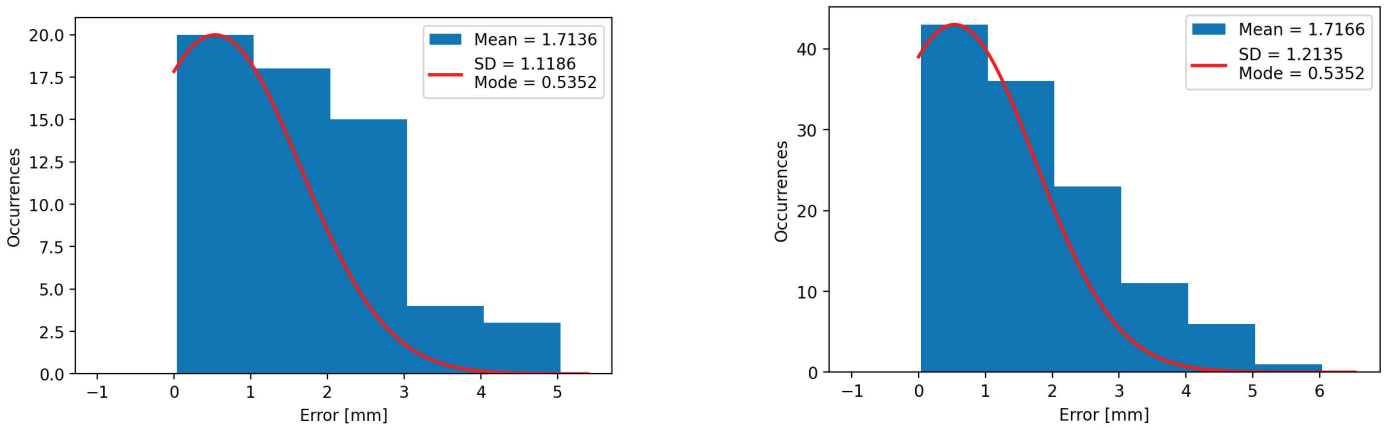
(**a**) (**b**)

Figure 11. (**a**) Error histogram on the X axis for natural light. (**b**) Error histogram on the X axis for artificial light.



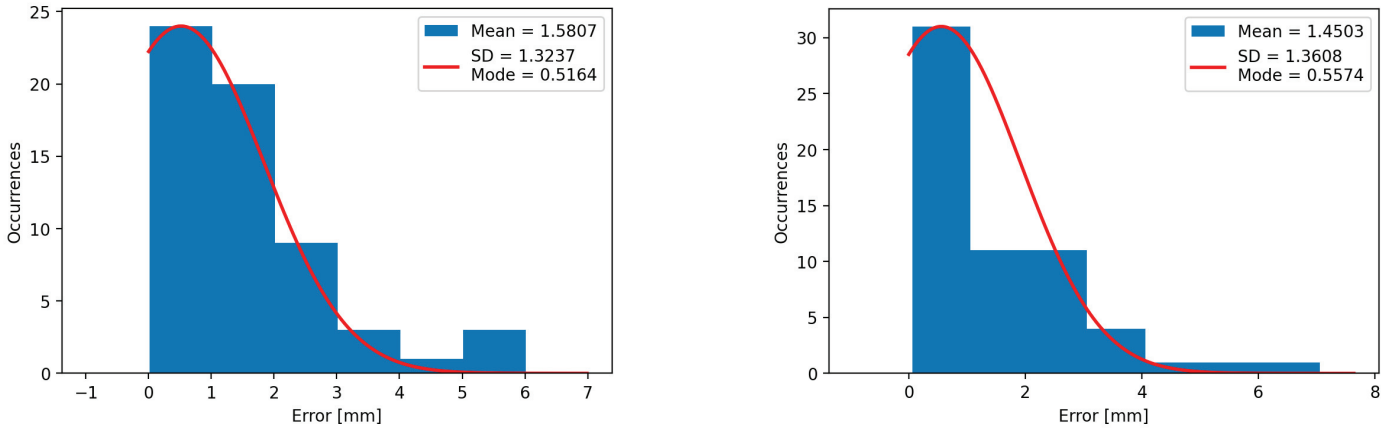
(**a**) (**b**)

Figure 12. (**a**) Error histogram on the X axis for natural light for the industrial software. (**b**) Error histogram on the X axis for artificial light for the industrial software.



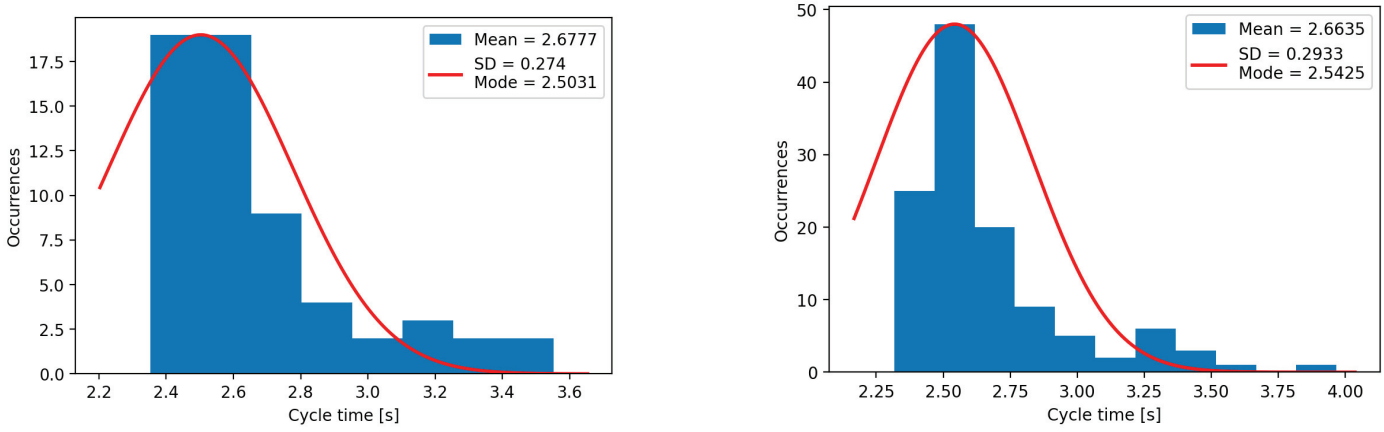
(**a**) (**b**)

Figure 13. (**a**) Error histogram on the Y axis for natural light. (**b**) Error histogram on the Y axis for artificial light.



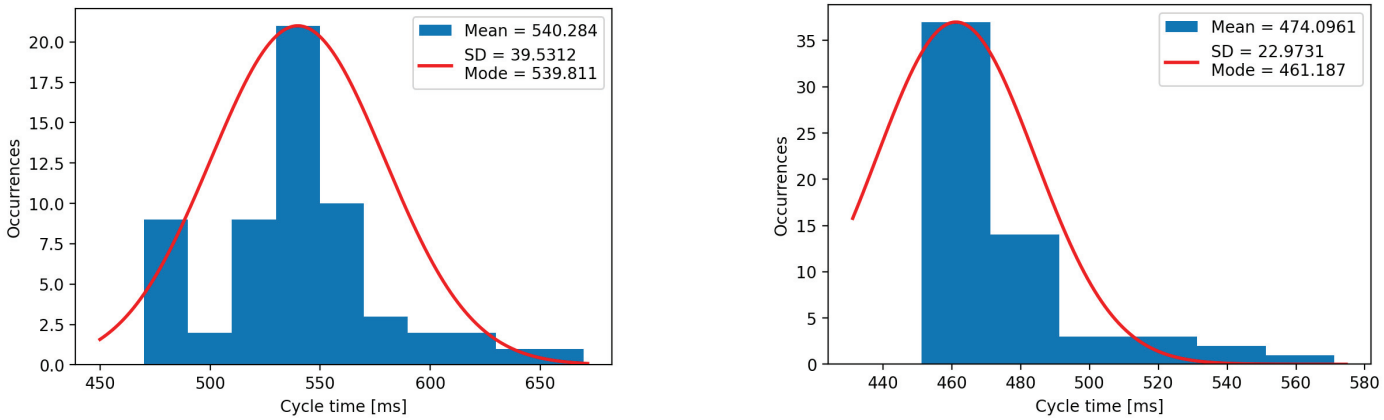
(**a**) (**b**)

Figure 14. (**a**) Error histogram on Y axis for natural light, for the industrial software. (**b**) Error histogram on Y axis for artificial light, for the industrial software.



(**a**) (**b**)

Figure 15. (**a**) Time histogram for natural light. (**b**) Time histogram for artificial light.



(**a**) (**b**)

Figure 16. (**a**) Time histogram for natural light for the industrial software. (**b**) Time histogram for artificial light for the industrial software.

5. Discussion

When using the real-world system, the images taken contain several added noises, such as the image background (the table) and the object's texture, shadows, reflections, and change in coloration due to the different light sources used, as compared to our previous work, where the images are taken inside a simulated environment. These noises decrease the detection confidence and increase the positioning error in the real environments and may pose a problem when performing robot sorting operations.

The best results are obtained when using natural light, under which the model was trained, and for the circular object due to its symmetry, but the robot system successfully picks the object in every case.

Compared with the industrial software, the confidence in detection is reduced when using the cloud service, but the position errors remain similar. Regarding the response time, the local solution is faster, as to be expected, but the cloud response is not slow either.

The response time varies with Internet speed, but it remains consistent under the same test conditions. The network used has a receive and transmit speed of 433 Mbps, the size of the images varies from 1.83 MB up to 2.04 MB, and the resolution is 2616×1960.

In comparison with classical identification methods, which take into account the shape and contour of the object, the presented method can be used successfully for identification in cases when the objects are a bit dirty (Figure 17) or they overlap (Figure 18).

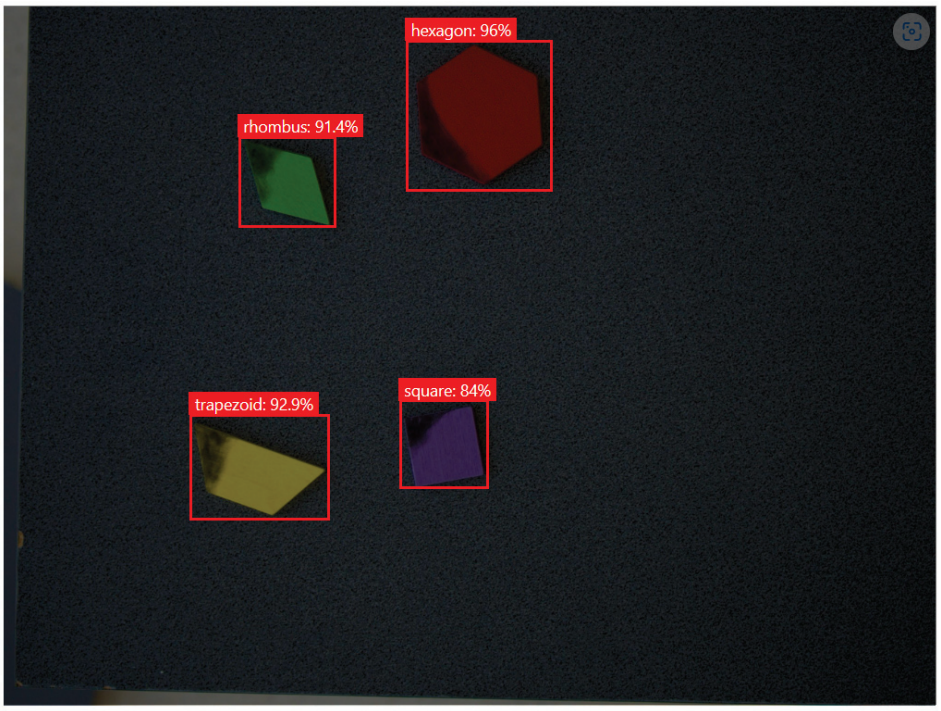


Figure 17. Detection results of dirty objects.

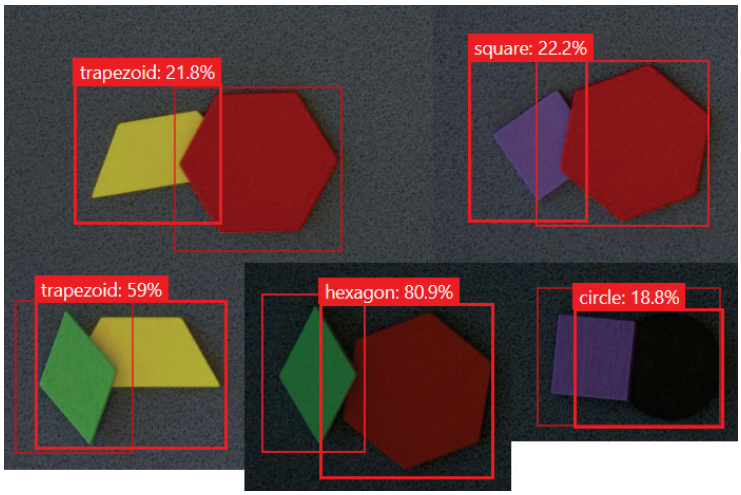
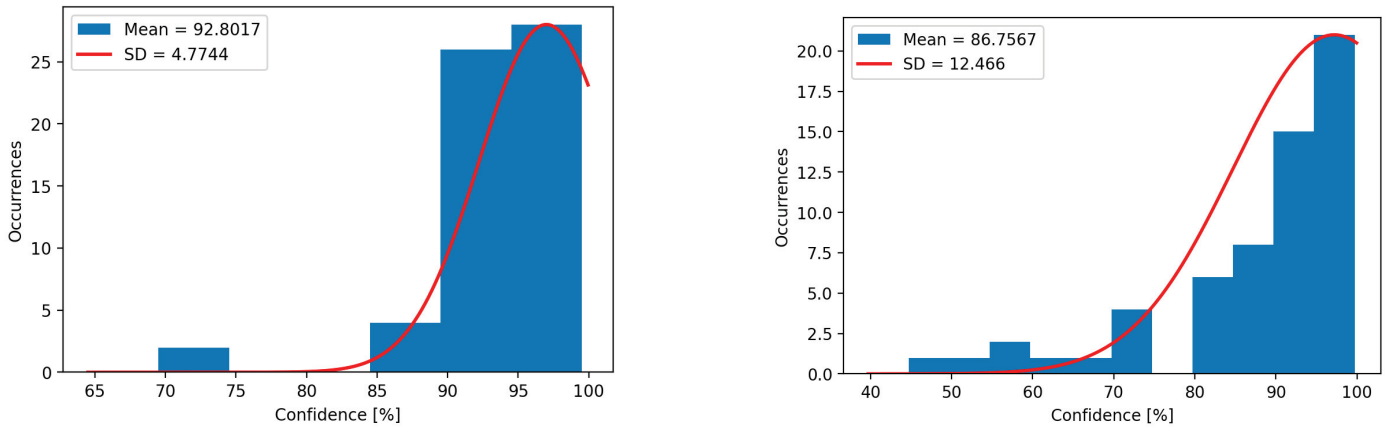


Figure 18. Detection results of overlapping objects.

The results are better when using the industrial software, trained for the same images, but when using the detection images to improve the cloud solution model, the detection results are improved. In this case, we used the detection images with natural light to retrain and improve the model, obtaining the results presented in Figure 19a for natural light, showing a 7.2% increase in detection confidence, and Figure 19b for artificial light, with a respectable 5.2% increase. Aside from the increase in confidence, a reduction in the standard deviation of 10.8% for natural light and 4.1% for artificial light is present.



(**a**) (**b**)

Figure 19. (**a**) Confidence histogram for natural light after retrain. (**b**) Confidence histogram for artificial light after retraining.

6. Conclusions

The cloud service obtained the best results in natural illumination conditions, with good prediction confidence and sufficiently few errors, ensuring object picking in every case. The implementation proved its overall connectivity consistency, while also demonstrating robustness, providing good results in the case of partially dirty objects.

The prediction confidence calculated by the Custom Vision and MVTec Merlic are not directly comparable since the methods of calculation of the confidence are not transparent for either of them. Per our observation, after making additional tests, the confidence of MVTec Merlic for cases of correct prediction tends to be closer to the extremes (either very high, which happens for most cases, or in specific cases, very low); we suspect that this behavior stems from the method used for calculating the confidence.

While the specialized software showed better and more consistent results overall, which is to be expected since it is a computer vision specifically made with industrial use in mind, the Custom Vision solution also performed satisfactorily. These results open additional application perspectives in the robotics field with quickly evolving AI technology and continuously increasing cloud capabilities.

It is worth mentioning that Custom Vision's lower limit of images for the training to take place is 25, and the results compared were obtained for a trained model where we used 40 training images. We observed that for further training of the Custom Vision model with new images of the objects, its prediction confidence kept increasing. The point after which the confidence is not further increased by adding new images is yet to be determined.

Despite the revolutionary potential of cloud computing, this technology also comes with certain limitations on aspects such as security, flexibility, or performance caused by Internet dependency. Further research needs to be conducted on this topic to ensure the safety of data while optimizing performance, which could benefit not only individual end-users, but also enterprises that are using these types of services.

In the future, a system that combines classical object identification, based on contours and edges, with NN-based detection, based on the two systems that are compared in this work, can be developed to account for all the detection scenarios, in terms of light, background, overlap, and dust, to obtain the best overall results.

Author Contributions: Conceptualization, I.-L.S. and A.M.; methodology, I.-L.S., A.M. and S.R.; software, I.-L.S., A.M. and I.L.; validation, I.-L.S., A.M. and S.R.; formal analysis, I.-L.S. and A.M.; investigation, I.-L.S. and A.M.; resources, I.-L.S., F.D.A. and S.R.; data curation, A.M.; writing—original

draft preparation, I.-L.S. and A.M.; writing—review and editing, I.-L.S., A.M., I.L., S.R., F.D.A., D.C.P. and I.S.S.; visualization, A.M.; supervision, S.R., D.C.P. and I.S.S.; project administration, A.M.; funding acquisition, F.D.A., I.-L.S. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Program for Research of the National Association of Technical Universities GNaC ARUT 2023, grant number 83/11.10.2023 (optDDNc).

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
CG	Correspondence Grouping
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
ELU	Exponential Linear Unit
GPU	Graphics Processing Unit
IoRT	Internet of Robotic Things
JSON	JavaScript Object Notation
NN	Neural Network
ONNX	Open Neural Network Exchange
ReLU	Rectified Linear Unit
SD	Standard Deviation
SELU	Scaled Exponential Linear Unit

References

1. Dawarka, V.; Bekaroo, G. Building and evaluating cloud robotic systems: A systematic review. *Robot. Comput.-Integr. Manuf.* **2022**, *73*, 102240. [CrossRef]
2. Zhang, H.; Zhang, L. Cloud Robotics Architecture: Trends and Challenges. In Proceedings of the 2019 IEEE International Conference on Service-Oriented System Engineering (SOSE), San Francisco, CA, USA, 4–9 April 2019; pp. 362–3625.
3. Stefan, I.L.; Mateescu, A.; Vlasceanu, I.M.; Popescu, D.C.; Sacala, I.S. Utilizing Cloud Solutions for Object Recognition in the Context of Industrial Robotics Sorting Tasks. In Proceedings of the 2024 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Craiova, Romania, 4–6 September 2024; pp. 1–6.
4. Fantozzi, I.C.; Santolamazza, A.; Loy, G.; Schiraldi, M.M. Digital Twins: Strategic Guide to Utilize Digital Twins to Improve Operational Efficiency in Industry 4.0. *Future Internet* **2025**, *17*, 41. [CrossRef]
5. Roumeliotis, K.I.; Tselikas, N.D.; Nasiopoulos, D.K. Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models. *Future Internet* **2025**, *17*, 28. [CrossRef]
6. Sahnoun, S.; Mnif, M.; Ghoul, B.; Jemal, M.; Fakhfakh, A.; Kanoun, O. Hybrid Solution Through Systematic Electrical Impedance Tomography Data Reduction and CNN Compression for Efficient Hand Gesture Recognition on Resource-Constrained IoT Devices. *Future Internet* **2025**, *17*, 89. [CrossRef]
7. Kamal, H.; Mashaly, M. Advanced Hybrid Transformer-CNN Deep Learning Model for Effective Intrusion Detection Systems with Class Imbalance Mitigation Using Resampling Techniques. *Future Internet* **2024**, *16*, 481. [CrossRef]
8. Karri, C.; Naidu, M.S.R. Deep Learning Algorithms for Secure Robot Face Recognition in Cloud Environments. In Proceedings of the 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Exeter, UK, 17–19 December 2020; pp. 1021–1028.
9. Zhang, Y.; Li, L.; Nicho, J.; Ripperger, M.; Fumagalli, A.; Veeraraghavan, M. Gilbreth 2.0: An Industrial Cloud Robotics Pick-and-Sort Application. In Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 38–45.

10. Jiang, J. The eye of artificial intelligence—Convolutional Neural Networks. *Appl. Comput. Eng.* **2024**, *76*, 273–279. [CrossRef]
11. Bryzgalov, P.; Maeda, T. Using Benchmarking and Regression Models for Predicting CNN Training Time on a GPU. In Proceedings of the PERMAVOST '24: 4th Workshop on Performance EngineeRing, Modelling, Analysis, and VisualizatiOn STrategy, Pisa, Italy, 3–4 June 2024; pp. 8–15. [CrossRef]
12. Li, X.; Li, R.; Zhao, Y.; Zhao, J. An improved model training method for residual convolutional neural networks in deep learning. *Multimed. Tools Appl.* **2021**, *80*, 6811–6821. [CrossRef]
13. Trigka, M.; Dritsas, E. Edge and Cloud Computing in Smart Cities. *Future Internet* **2025**, *17*, 118. [CrossRef]
14. Souza, D.; Iwashima, G.; Farias da Costa, V.C.; Barbosa, C.E.; de Souza, J.M.; Zimbrão, G. Architectural Trends in Collaborative Computing: Approaches in the Internet of Everything Era. *Future Internet* **2024**, *16*, 445. [CrossRef]
15. Maenhaut, P.-J.; Moens, H.; Volckaert, B.; Ongenae, V.; De Turck, F. Resource Allocation in the Cloud: From Simulation to Experimental Validation. In Proceedings of the 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), Honolulu, HI, USA, 25–30 June 2017; pp. 701–704. [CrossRef]
16. Saeed, H.A.; Al-Janabi, S.T.F.; Yassen, E.T.; Aldhaibani, O.A. Survey on Secure Scientific Workflow Scheduling in Cloud Environments. *Future Internet* **2025**, *17*, 51. [CrossRef]
17. Jansi Sophia Mary, C.; Mahalakshmi, K.; Senthilkumar, B. Deep Dive On Various Security Challenges, Threats And Attacks Over The Cloud Security. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 17–18 March 2023; pp. 2089–2094. [CrossRef]
18. Deepika; Kumar, R.; Dalip. Security Enabled Framework to Access Information in Cloud Environment. In Proceedings of the 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 26–27 May 2022; pp. 578–582. [CrossRef]
19. Atanasov, I.; Dimitrova, D.; Pencheva, E.; Trifonov, V. Railway Cloud Resource Management as a Service. *Future Internet* **2025**, *17*, 192. [CrossRef]
20. Ji, P.; Chu, H.; Chi, J.; Jiang, J. A Resource Elastic Scheduling Algorithm of Service Platform for Cloud Robotics. In Proceedings of the 2020 Chinese Control and Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 4340–4344.
21. Raheel, F.; Raheel, L.; Mehmood, H.; Kadri, M.B.; Khan, U.S. Internet of Robots: An Open-source Framework for Cloud-Based Global Remote Control and Monitoring of Robots. In Proceedings of the 2024 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, 18–19 December 2024; pp. 1–6.
22. Chou, J.; Chung, W.-C. Cloud Computing and High Performance Computing (HPC) Advances for Next Generation Internet. *Future Internet* **2024**, *16*, 465. [CrossRef]
23. Basler ace 2 Ra2A2590-22gcBAS. Available online: https://www.baslerweb.com/en/shop/a2a2590-22gcbas/ (accessed on 16 March 2025).
24. Basler Lens C125-1620-5M f16mm. Available online: https://www.baslerweb.com/en/shop/lens-c125-1620-5m-f16mm/ (accessed on 16 March 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Future Internet Editorial Office
E-mail: futureinternet@mdpi.com
www.mdpi.com/journal/futureinternet



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-8870-2