

Article

A Novel Enterprise AI Classification Framework for Business Transformation: A Structured Literature Review and Integration of AI Types and Autonomy Levels

Nusi Borovac ¹  and Vladimir Stantchev ^{2,*} 
¹ Kenza GmbH, Unter den Linden 40, 10117 Berlin, Germany; nusi.borovac@kenza.io

² TEAC, Institute of Information Systems, SRH University Heidelberg, 69123 Heidelberg, Germany

* Correspondence: stantchev@computer.org

Abstract

Enterprise investment in artificial intelligence has reached an unprecedented scale across national, regional, and organisational levels of the economy, yet transformation outcomes remain highly variable. Recent global research indicates that the majority of enterprise AI investments produce no measurable profit-and-loss impact, with only a minority of organisations extracting material enterprise-level value. Without an integrated classification framework that allows enterprises to deploy, govern, and create value from the full range of AI technologies and autonomy levels across enterprise functions, the risks of misallocated investment, fragmented deployment, and unrealised return remain difficult to mitigate. A structured literature review on AI classification reveals that existing frameworks do not integrate AI technology types, operational autonomy levels, and concrete enterprise applications in a single classification structure that can support enterprise AI-driven transformation decisions. To address this gap, this paper proposes the Enterprise AI Classification Framework, a novel classification model that integrates six AI technology types with six autonomy levels in a 6×6 matrix of 36 combinations, each corresponding to concrete business applications across enterprise functions. A computational pilot study ($n = 20$ cases for case-based coding validation and 12 LLM-simulated personas $\times$ 30 vignettes for role-specific hypothesis pre-specification) provides preliminary evidence of substantially higher classification coverage than four baseline frameworks and pre-specifies role-divergence hypotheses for a planned empirical validation, with all design choices locked in an OSF pre-registration; reliability point estimates exceed the pre-specified $\alpha \geq 0.70$ threshold at the pilot scale, with bootstrapped confidence intervals expected to tighten under the full-corpus run in the follow-up study. An interview-based empirical validation with approximately 250 decision-makers and practitioners across 20 industries is in preparation.



Academic Editor: Peiying Zhang,
Athanasios V. Vasilakos

Received: 2 June 2026

Revised: 22 June 2026

Accepted: 26 June 2026

Published: 1 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: artificial intelligence; AI classification framework; enterprise transformation; AI autonomy levels; AI technology types; digital transformation; enterprise AI

1. Introduction

1.1. The Business Risk of AI Transformation Without Adequate Classification

The use of artificial intelligence (AI) is a defining strategic question of the current decade at every level of economic organisation. National governments articulate AI strategies as instruments of long-term industrial policy. China's 15th Five-Year Plan for National

Economic and Social Development (2026–2030), adopted by the National People’s Congress on 12 March 2026, places AI at the centre of national strategy and calls for progress on the comprehensive “AI+” action plan, which was articulated by the State Council in 2024 and 2025 and is carried forward across science and technology, industrial development, consumption, public services, governance, and international cooperation [1]. Regional bodies are deploying capital and regulatory architecture at scale. The European Union’s Artificial Intelligence Act, Regulation (EU) 2024/1689, was adopted by the European Parliament and the Council on 13 June 2024 and published in the *Official Journal of the European Union* on 12 July 2024 as the first comprehensive horizontal binding legal framework on AI [2]. The InvestAI initiative, launched by the European Commission at the AI Action Summit in Paris on 11 February 2025, mobilises €200 billion of investment in AI, including a new European fund of €20 billion for AI gigafactories [3]. InvestAI is complemented by the AI Continent Action Plan, COM(2025) 165 of 9 April 2025 [4], and the Apply AI Strategy, COM(2025) 723 final of 8 October 2025 [5]. Member states have supplemented this architecture with national commitments, including the announcement by the president of the French Republic at the Paris AI Action Summit on 10 February 2025 of €109 billion in private, foreign and French investment commitments in AI in France [6]. The classification and deployment of AI is now a universal strategic question, and the answer at each level shapes the economic trajectory of nations, regions, and organisations of every size.

At the enterprise level, this strategic investment translates into specific business risks. A board commissioning AI transformation without a classification framework faces misallocated investment in AI technologies poorly matched to their applications, fragmented deployment in which AI systems proliferate in isolated pilots without enterprise coherence, misaligned expectations between board, functional leaders, and technical teams, regulatory exposure under the EU AI Act and comparable frameworks that require categorised and auditable AI deployment, competitive disadvantage relative to peers who classify and deploy AI systematically, and transformation failure in which AI projects do not survive beyond the pilot stage. These risks are not theoretical. Recent practitioner research—institutional grey literature and consultancy surveys rather than peer-reviewed empirical work, but consistently triangulating the same underlying market signal across independent data sources—documents the gap between enterprise AI investment and measurable financial impact. The MIT Project NANDA report *The GenAI Divide: State of AI in Business 2025* [7] reports that after \$30–40 billion of enterprise spend, approximately 95% of organisations are seeing no measurable profit-and-loss impact from generative AI, with only 5% of integrated AI pilots extracting millions in value. Gartner projected that over 40% of agentic AI projects will be cancelled by the end of 2027, citing escalating costs, unclear business value, and inadequate risk controls [8]. The McKinsey Global Survey on the state of AI 2025, with 1993 respondents across 105 nations, found that nearly two-thirds of organisations have not yet begun scaling AI across the enterprise, and that even among the minority reporting any EBIT impact from AI, most attribute less than 5% of EBIT to its use [9]. The MIT CISR *Building Enterprise AI Maturity* briefing, drawing on the October 2022 MIT CISR Future Ready Survey of 721 companies, found that enterprises in the first two stages of its four-stage AI maturity model performed below industry average on financial measures, while enterprises in stages three and four performed well above industry average [10]. A systematic literature review of artificial intelligence maturity models found a substantial proliferation of competing models, the majority of which are domain-specific, and concluded that no unified framework spans the full set of AI technology types [11]. The peer-reviewed evidence reaches the same conclusion from several directions. Enholm et al. [12], reviewing the literature on AI and business value, trace value to organisational and deployment complementarities, not to buying the technology;

Mikalef and Gupta [13] show empirically that the performance gains accrue to firms with a measurable AI capability built from tangible, human, and intangible resources, not to firms that merely spend on AI. The productivity returns are real but unevenly shared: Noy and Zhang [14] measure a 40% cut in task time and an 18% gain in output quality on a controlled writing task, and Brynjolfsson, Li, and Raymond [15] measure a 15% average lift in customer-support productivity that rises to 34% for novices and is negligible for experts. The exposure is economy-wide, not sectoral—Eloundou et al. [16] estimate that large language models touch a substantial share of tasks across most occupations—which is exactly why the deployment question this paper addresses is a general one. Collins et al. [17], surveying AI in information systems research, name the missing piece directly: integrative frameworks that connect AI capabilities to how the enterprise actually deploys them. Investing in AI does not by itself produce transformation; the capacity to classify, structure, and deploy AI across the enterprise is what distinguishes the organisations that realise the return on AI investment from those that do not.

A central cause of these risks is a knowledge gap. Business leaders commissioning AI transformation face questions their existing frameworks cannot answer: which types of AI technology exist, at what level of autonomy each could operate, and what concrete applications of these combinations look like across enterprise functional areas. The academic literature on AI classification, surveyed in Section 2 of this paper, is substantial but largely written for technologists, researchers, and regulators. This paper proposes a classification framework at the enterprise level, where the question is most immediately operational. A fourth adjacent area, AI governance, encompassing risk management, accountability structures, ethical oversight, and compliance frameworks, is complementary to classification but methodologically distinct, and is not integrated into the present study. Each of these works organises a facet that classification deliberately leaves alone. Newman [18] builds a taxonomy of trustworthiness that ties properties such as reliability, safety, and accountability to the stages of the AI lifecycle and to risk management practice. Abercrombie et al. [19] classify systems by the harms they produce rather than by their technical type. Bagehorn et al. [20] map AI risks to the resources and tooling that mitigate them. One level up, the systematic reviews of Mäntymäki et al. [21] and Birkstedt et al. [22] consolidate organisational AI governance and report that the field does not yet share a settled definition. The division of labour is clean: the present framework classifies what an AI deployment *is*; this literature classifies how it should be *controlled*. The two layers interoperate, and we keep them apart on purpose.

1.2. Research Questions

Three research questions, sequenced as survey, gap, and proposal, frame this paper.

- RQ1.** Which AI classification frameworks and approaches are available for AI-driven enterprise transformation?
- RQ2.** What are the current research gaps in AI classification frameworks for AI-driven enterprise transformation?
- RQ3.** Which new AI classification framework can address the identified gaps to support AI-driven enterprise transformation?

RQ1 is addressed in Section 2 (the structured literature review) and consolidated in Section 4.1. RQ2 is addressed in Section 4.2, which presents the gap analysis. RQ3 is addressed in Section 4.3, which proposes the Enterprise AI Classification Framework, and is reinforced in Section 4.4 through a structured comparison with prior work.

1.3. Contribution and Structure of This Paper

This paper proposes the Enterprise AI Classification Framework: a classification structure for AI across the enterprise, consisting of a two-dimensional matrix that combines six AI technology types with six autonomy levels. The thirty-six-combination matrix is populated with concrete business application examples drawn from enterprise functional contexts. The framework makes three contributions. The theoretical contribution is the integration of AI technology types and autonomy levels in a single classification structure at the enterprise deployment unit of analysis. The closest historical precedent builds on the levels-of-automation scale of Sheridan and Verplank [23]; Parasuraman, Sheridan, and Wickens [24] pair a types-like axis (cognitive processing stages: information acquisition, information analysis, decision and action selection, action implementation) with those autonomy levels within a single automated system; the present framework re-grounds an integration of the same general form at the enterprise unit of analysis, replacing cognitive-processing-stage types with AI-as-deployed technology types and populating the matrix with concrete enterprise-function applications. The novelty of the present framework consists precisely in this re-grounding and the cross-stream synthesis it enables; an architecture of this form is not present across the three research streams reviewed. The methodological contribution is the translation of technically grounded classification into a form designed for business use, populated with applications business leaders recognise. The practical contribution is an operational framework for enterprises commissioning AI transformation, providing boards, functional leaders, and technical teams with a shared taxonomy through which AI deployment can be described, located, and discussed in consistent terms.

The remainder of this paper is organised as follows. Section 2 reviews the existing literature across three research streams. Section 3 sets out the methodology of the review and the framework construction approach. Section 4 presents the principal findings: a consolidated summary of the literature reviewed (Section 4.1), the gap analysis (Section 4.2), the Enterprise AI Classification Framework (Section 4.3), and a structured comparison with prior frameworks (Section 4.4). Section 5 reports a computational pilot study evaluating the framework. Section 6 discusses the contributions and limitations. Section 7 concludes this work.

2. Structured Literature Review

The classification of AI systems for enterprise transformation can be drawn from different research streams: automation and autonomy foundations, AI taxonomies and capability classifications, and enterprise and generative AI frameworks. Each stream has produced frameworks with their own internal logic and methodological conventions. None has combined the elements of AI technology types, autonomy levels, and enterprise transformation classification in a single integrated framework. This section reviews each stream in turn, describing the contributions of the cited works and identifying the structural absences they reveal in the domain of AI-driven enterprise transformation. The cross-stream synthesis and the gap statement that motivates the framework are presented in Section 4.

2.1. Stream A: Automation and Autonomy Foundations

Stream A originates in teleoperation and human-machine systems research, where the central question is how authority is distributed between human operators and computer control. The stream has developed in three phases: a foundational phase establishing a unidimensional levels-of-autonomy scale; a second phase introducing a two-dimensional model that separates types of automation from levels of autonomy; and a third phase, contemporary with foundation models and agentic AI, that produces multi-dimensional classification frameworks.

Sheridan and Verplank [23] proposed a ten-level scale of levels of automation in human–computer decision-making in their 1978 MIT Man–Machine Systems Laboratory technical report on undersea teleoperators. The report orders the levels from a level at which the human does the whole job up to a level at which the computer does the whole job. The scale has shaped subsequent work across automotive, aerospace, healthcare robotics, and human–computer interaction research. The unit of analysis is the individual automated system; classification of AI technology types and enterprise deployment contexts are outside the scope of the report.

Parasuraman, Sheridan, and Wickens [24] proposed that automation can be applied to four classes of functions, which they label types: information acquisition, information analysis, decision and action selection, and action implementation. Each type can be automated to differing degrees. The ten-point scale they propose, based on the previously proposed scale of Sheridan and Verplank [23], refers mainly to automation of decision and action selection; the authors note that the scale can be applied with some modification to information acquisition, information analysis, and action implementation, although the number of levels will differ between the stages. The framework is intended to guide engineering choices about where and at what level to introduce automation within an individual automated system; classification at the enterprise level is outside the scope of this article.

Dellermann, Ebel, Söllner, and Leimeister [25] proposed the concept of hybrid intelligence, defined as the ability to achieve complex goals by combining human and artificial intelligence, thereby reaching superior results to those each of them could have accomplished separately, and continuously improve by learning from each other. The framework is configurational rather than level-based; it identifies design dimensions along which human and machine contributions can be distributed for a given task, rather than proposing an ordinal scale of autonomy. The framework is designed for individual hybrid systems; it does not offer an enterprise classification or a mapping across AI technology types.

Morris et al. [26] proposed the Levels of AGI framework. The framework introduces levels of AGI performance, generality, and autonomy as a common language to compare models, assess risks, and measure progress along the path to artificial general intelligence. Performance and generality are presented in a matrixed taxonomy, and autonomy is presented separately as a six-level scale of human–AI interaction paradigms. The framework is centred on the AI system and on the path to AGI rather than on enterprise deployment.

Feng, McDonald, and Zhang [27] defined five levels of escalating agent autonomy characterised by the roles a user can take when interacting with an agent: operator, collaborator, consultant, approver, and observer. The framework treats autonomy as a deliberate design decision separate from an agent’s capability and operational environment, and is intended to inform agent governance and design. Like the AGI framework, the unit of analysis is the individual AI agent; classification across AI technology types or across an enterprise is not provided.

Porter et al. [28] proposed the INSYTE classification framework. INSYTE considers the essential characteristics of an AI system across eight dimensions grouped into four categories: system design (underspecification and adaptiveness); system functionality (breadth of functionality and depth of functionality); operating environment (environmental diversity and environmental dynamism); and operational independence (independence from intervention and independence from oversight). Each dimension is scored on a scale from Level 0 to Level 5, and the system’s combination of scores is depicted on an eight-axis radar chart. The framework supports classification of AI systems ranging from traditional rule-based systems to embodied and agentic AI and is aligned with the OECD definition of a deployed AI system. Its focus is the characterisation of individual AI

systems across multiple dimensions; enterprise transformation application is outside the framework's purpose.

SAE International published the standard J3016 [29] as the recognised international taxonomy and definition for terms related to driving-automation systems for on-road motor vehicles. The standard distinguishes six levels of driving automation from Level 0 (no driving automation) to Level 5 (full driving automation), defined by the division of dynamic driving task between the human driver and the automated driving system across longitudinal and lateral vehicle motion, object and event detection and response, and fallback. Although developed for on-road motor vehicles, the SAE J3016 six-level scheme has become the dominant lay reference point for “levels of autonomy” across general AI discourse, regulatory documents, and the management literature, and consequently it serves as a cognitive anchor for the autonomy axis presented in Section 4.3.2. The unit of analysis in J3016 is, like the rest of Stream A, the individual automated system; the standard does not extend to enterprise classification, to AI technology types beyond automated driving, or to deployment-level division of authority across heterogeneous enterprise functions.

Across the trajectory from a unidimensional levels-of-autonomy scale in 1978 to SAE J3016 (2021) and the eight-dimension INSYTE framework (2025), the unit of analysis in Stream A has remained the individual AI system. The stream provides a detailed analytical framework for characterising the autonomy and structural properties of an AI system but does not extend to the classification of AI across an enterprise.

2.2. Stream B: AI Taxonomies and Capability Classifications

Stream B comprises classifications of AI as a class of technology, developed across three institutional sources: European and United States policy research bodies, management and information systems researchers, and labour economists. The stream's unifying question is what AI is, what it does, and which tasks it is suited to; the unit of analysis is the technology or the task rather than the deployed system.

The EIT Cross-KIC taxonomy of the European AI ecosystem [30], produced under the Cross-KIC activity “Innovation Impact Artificial Intelligence” with EIT Climate-KIC as the lead work package, surveyed thirty-five existing AI classification frameworks and consolidated their content along eight observed dimensions of comparison. The taxonomy itself organises the European AI landscape across five dimensions: industries, enterprise functions, locations, AI capabilities, and enabling technology types. Within the AI capabilities dimension, the report identifies eight broader capability categories: computer vision, computer audition, computer linguistics, robotics, forecasting, discovery, planning, and creation. The framework is positioned as a reference for European AI ecosystem mapping and policy coordination. Its capability axis is single-dimensional and is not paired with autonomy levels.

Samoili et al. [31] published *AI Watch. Defining Artificial Intelligence 2.0: Towards an Operational Definition and Taxonomy for the AI Landscape* as a Joint Research Centre technical report. The taxonomy organises AI into a set of core domains (reasoning, planning, learning, communication, and perception) and transversal domains (integration and interaction, services, and ethics and philosophy), each with associated subdomains. The taxonomy was developed to support the European Commission AI Watch initiative, which monitors AI policy and adoption across the European Union. Its purpose is policy monitoring and cross-country comparison; the taxonomy does not pair domains with autonomy levels or with an enterprise classification.

Theofanos, Choong, and Jensen [32] published the *AI Use Taxonomy: A Human-Centered Approach* as NIST AI 200-1. The taxonomy comprises sixteen human-AI activities

defined in Table 1 of the report: content creation, content synthesis, decision making, detection, digital assistance, discovery, image analysis, information retrieval/search, monitoring, performance improvement, personalization, prediction, process automation, recommendation, robotic automation, and vehicular automation. The taxonomy is human-centred, positioning the user and the use context at the centre of the classification, and supports the NIST AI Risk Management Framework. The taxonomy describes AI activities along a single axis; pairing with autonomy levels is outside its scope.

Table 1. Stream-specific search strings.

Stream	Search String (Boolean)
A	("levels of autonomy" OR "levels of automation" OR "autonomy framework" OR "human–AI authority" OR "levels of AGI") AND ("classification" OR "taxonomy" OR "framework")
B	("AI taxonomy" OR "AI classification" OR "AI capability taxonomy" OR "AI use taxonomy") AND ("operational definition" OR "policy" OR "capability categories")
C	("enterprise AI" OR "AI maturity" OR "generative AI taxonomy" OR "agentic AI" OR "multi-agent system") AND ("framework" OR "classification" OR "maturity model" OR "business transformation")

Davenport and Ronanki [33] proposed a management-oriented classification of cognitive technologies organised by the type of business task each addresses: process automation, cognitive insight, and cognitive engagement. Davenport’s subsequent monograph *The AI Advantage* [34] elaborates the taxonomy into a practitioner reference framework for enterprise AI deployment. The classification is influential in management practice and in the executive-education literature, but it organises AI by business task addressed rather than by AI technology type or by autonomy level, and the three categories are coarser than the categories required for an enterprise to identify and procure AI as a discrete budgetary unit. The taxonomy does not pair its categories with autonomy levels.

Brynjolfsson and Mitchell [35] proposed a Suitability for Machine Learning (SML) rubric assessing the degree to which a given task is amenable to current machine learning techniques. The rubric scores tasks on eight criteria including the availability of large training data, the well-defined nature of input–output mappings, the absence of long chains of logical reasoning, and the tolerance for errors. The companion task-based study by Brynjolfsson, Mitchell, and Rock [36] applies the SML rubric across occupations and tasks in the United States economy. The framework is task-classificatory rather than technology-classificatory: it asks how machine-learning-suited a task is, not what type of AI technology a deployment is or at what autonomy level it operates. The framework is widely cited in the management and labour economics literatures as a complement to, rather than substitute for, technology-type taxonomies.

Stream B provides well-developed classifications of what AI is, what it does, and which tasks it is suited to. Its unit of analysis is AI as a class of technology rather than the individual system or the enterprise. The stream provides a structured answer to the question of what categories of AI exist but does not pair those categories with autonomy levels or with a classification of enterprise deployment contexts.

2.3. Stream C: Enterprise and Generative AI Frameworks

Stream C has developed across management research, information systems scholarship, computational science, and consulting practice. The stream encompasses structural taxonomies, capability hierarchies, agentic architectures, and adoption-stage maturity models that organise AI at the level of the enterprise. Its unit of analysis is the enterprise as a deployer of AI.

Herrmann [37] proposed a unified framework for artificial intelligence in enterprise applications, based on a systematic review and science mapping of 26,143 publications. The framework is depicted as an Euler diagram in which two structural dimensions are mapped: a vertical dimension reflecting a hierarchy of analytics together with computer hardware and planning/optimisation, and a horizontal dimension that maps decision support system types separately from AI and identifies overlap with three AI paradigms (symbolic, algorithmic, and metaheuristic) together with supporting information- and communication-technology infrastructure. The framework is directed at enterprise AI but does not pair structural categories with autonomy levels.

Bashir [38] proposed the Strategic Enterprise Artificial Intelligence (SEAI) conceptual hierarchical framework, distinguishing four levels: disconnected enterprise AI, Linked enterprise AI, strategist enterprise AI, and integrative enterprise AI. The framework characterises the strategic linkage of enterprise AI to business strategy at four progressively integrated stages. It does not pair the structural categories with autonomy levels and does not offer concrete business application examples mapped across AI technology types.

Stein [39] proposed a taxonomy of generative AI use cases in business contexts. The framework synthesises five perspectives drawn from academic and industry literature (application context, value creation, strategic alignment, technical autonomy, and data governance) and organises them through a four-layer complexity ladder running from (A) work assistants, (B) automated code generation, and (C) system-integrated text generation, to (D) tool use. The taxonomy is mapped across two application contexts: internal employee empowerment and external customer experience enhancement. Its scope is generative AI specifically rather than the full range of AI technology types.

Sapkota, Roumeliotis, and Karkee [40] proposed a conceptual taxonomy distinguishing AI agents from agentic AI. The taxonomy characterises AI agents as modular systems driven by large language models for task-specific automation, advanced through tool integration, prompt engineering, and reasoning enhancements; in contrast, it characterises agentic AI as a paradigm marked by multi-agent collaboration, dynamic task decomposition, persistent memory, and coordinated autonomy. The taxonomy provides comparative analysis across the two paradigms and clarifies vocabulary in a literature in which terms such as agent, agentic, and autonomous are used inconsistently.

Panigrahy [41] proposed the Multi-Agentic AI Systems framework, organising enterprise-grade multi-agent systems into four layers: an Environment Layer that integrates the framework with organisational systems through standardised application programming interfaces; an Agent Layer containing specialised agents categorised by function as perception, cognition, action, and coordination agents; a Knowledge Layer for information management, ontologies, and machine learning repositories; and a Control and Management Layer for governance, policy enforcement, and monitoring. The framework addresses the architectural structure of multi-agent enterprise AI but does not provide a classification across AI technology types beyond agentic AI.

Arunkumar, Gangadharan, and Buyya [42] published a unified taxonomy of agentic AI architectures on arXiv, decomposing agents into six functional categories—perception, brain, planning, action, tool use, and collaboration—and reviewing the transition from single-loop agents to hierarchical multi-agent systems. The autonomy treated by Arunkumar et al. [42] is that of the AI agent itself, considered at the system-architecture level; this differs from the deployment-level division of authority used in the present paper, which describes how decision rights are split between human and AI for a given enterprise deployment regardless of the system's underlying capability.

Sadiq et al. [11] published a systematic literature review of AI maturity models, drawing on eight academic sources and finding that the most frequent number of maturity levels

is five and that maturity grids and continuous representations with five levels are currently trending. Fornasiero et al. [43] extend this line empirically, proposing and applying an AI and big-data maturity model across a sample of European process-industry firms and reporting differentiated maturity across operational dimensions. Weill, Woerner, and Sebastian [10] developed the Building Enterprise AI Maturity model, drawing on a 2022 survey of 721 companies and interviews with sixteen executives at nine enterprises. The model defines four stages of enterprise AI maturity—experiment and prepare, build pilots and capabilities, develop AI ways of working, and become AI future ready—each associated with distinct financial performance characteristics. Both maturity-model frameworks present enterprise AI maturity as a single-axis staged progression and do not specify which AI technology types operate at which autonomy levels.

Iansiti and Lakhani [44], in *Competing in the Age of AI*, proposed an enterprise-level model of the “AI factory” in which a firm’s data pipeline, algorithm development, experimentation platform, and software infrastructure jointly constitute a learning operating model that scales without the diminishing returns of traditional operating models. The framework characterises the structural shift from process-centric to AI-centric enterprise operation and is a widely cited reference point in management literature on enterprise AI strategy. It does not propose a classification of AI technology types or of autonomy levels, and the framework’s categories operate at the level of the firm-wide operating model rather than the individual deployment.

Acemoglu and Restrepo [45] articulate a typology distinguishing automation that displaces labour from technologies that reinstate labour through new tasks. The typology informs the macroeconomic literature on AI and employment and provides categories for analysing the labour-market consequences of automation, but it is not a classification of AI systems by AI technology type or autonomy level for enterprise deployment.

Beyond academic frameworks, two standards bodies and a small group of practitioner-research organisations contribute reference structures relevant to enterprise AI classification. ISO/IEC 42001:2023 [46] specifies requirements for an AI management system at the organisation level, prescribing controls covering AI policy, AI system lifecycle, data management, third-party relationships, and continual improvement, but does not propose a classification of AI by AI technology type or autonomy level. The NIST AI Risk Management Framework [47] provides a voluntary framework for managing risks across the AI lifecycle through four core functions (govern, map, measure, and manage) and complements the NIST AI 200-1 use taxonomy [32] as a paired use-and-risk pair, but does not pair AI activity categories with autonomy levels. Major practitioner research organisations publish recurring AI frameworks across multiple scales of analysis. The Boston Consulting Group produces both a country-level AI Maturity Matrix scoring seventy-three economies across six ASPIRE indicators (ambition, skills, policy and regulation, investment, research and innovation, and ecosystem) [48] and an enterprise-facing AI Radar tracking C-suite adoption [49]; Deloitte’s *State of Generative AI in the Enterprise* [50] (quarterly survey series, wave four fielded July–September 2024, $n = 2773$ respondents) and Accenture’s *Technology Vision* [51] (annual trends report, twenty-fourth year, 2024) provide complementary practitioner views. None extends the multi-dimensional structural rigour of the country-level work to enterprise-facing AI deployment classification with formal type and level axes; the gap this paper addresses is precisely that extension.

One adjacent literature reference studies the enterprise as a deployer of AI through the lens of value and capability rather than classification, and it bounds the present contribution from the side. Enholm et al. [12] review AI and business value and trace value to organisational complementarities; Mikalef and Gupta [13] operationalise an AI capability construct and tie it to creativity and firm performance; Collins et al. [17] map the AI information

systems agenda and flag the scarcity of integrative deployment frameworks; Kanbach et al. [52] read generative AI through business-model innovation and derive propositions on how it reshapes value creation and capture across software, healthcare, and financial services. The limit is the same one in every case: this work establishes *whether* and *how much* value enterprise AI creates, but not *which* type of AI runs at *which* level of autonomy. That descriptive question is the one the present framework answers, and a value model can be built on top of it.

Stream C provides structured taxonomies, agentic architectures, staged maturity models, and standards-derived management systems for enterprise AI. The stream does not integrate AI technology types with autonomy levels and does not map concrete business application examples simultaneously across both dimensions for enterprise transformation.

2.4. Section Signpost

The three streams reviewed above together constitute a body of work on AI classification, each contributing a distinct perspective at a distinct unit of analysis. Section 4 returns to this body of work to consolidate the principal findings of the review (Section 4.1), to articulate the structural gap that motivates the framework proposed in this paper (Section 4.2), to introduce the Enterprise AI Classification Framework as the response to that gap (Section 4.3), and to position the framework against the prior work reviewed here through a structured comparison (Section 4.4).

3. Methodology

This section sets out the methodology in three parts: the scope and selection criteria applied to the literature, the construction approach used to develop the Enterprise AI Classification Framework, and the forward plan for empirical validation. The present contribution is the framework, the literature analysis that motivates it, and a computational pilot study reported in Section 5; an interview-based empirical validation is forthcoming. The research programme as a whole follows a review-then-empirical architecture—a structured literature review and integrated framework, followed by an empirical validation study—whose methodological lineage is set out in Section 3.2.1.

3.1. Review Scope and Selection Criteria

The literature review reported in Section 2 was conducted as a structured literature review following the PRISMA 2020 reporting items [53] adapted to a non-clinical, conceptual review. The review does not claim exhaustive coverage of every relevant work; rather, it identifies, screens, and includes the contemporary classification frameworks most directly relevant to the structural gap that motivates the present paper. This scoping decision is consistent with the review-then-build approach followed in the authors' prior work, in which a comparable structured review identified a defined set of adoption models and integrated theories prior to proposing a new integrated model (Section 3.2.1).

Information Sources and Search Dates.

The information-source list comprises six electronic resources, which we grouped by type rather than treat as interchangeable databases: two multidisciplinary citation indexes (Web of Science Core Collection and Scopus), three publisher full-text platforms (Elsevier ScienceDirect, ACM Digital Library, and IEEE Xplore), and one academic search engine that aggregates across publishers without applying an editorial quality filter (Google Scholar, used only for citation tracing and coverage checks rather than as a primary identification source). These are supplemented by structured search of the institutional repositories of European and United States policy research bodies (European Institute of Innovation and Technology, European Commission Joint Research Centre, National Institute of Standards

and Technology, MIT Center for Information Systems Research, and SAE International). The search window covered publications from 1978 (Sheridan and Verplank, the foundational anchor work for Stream A) to the search closing date of 1 May 2026. The identification counts reported in Figure 1 were obtained via the OpenAlex API (<https://api.openalex.org>, accessed on 25 June 2026). OpenAlex is a free, openly licensed bibliographic database covering approximately 250 million scholarly works across all major venues. It is used here as a comprehensive substitute for the closed databases that require institutional API credentials. The closed-database queries are reserved for the full-corpus run in the follow-up study; the included-works set is robust to the substitution, since the included works are identified through structured selection rather than through database hit counts alone. On 1 May 2026 with the Boolean phrase set combined using OpenAlex's `title_and_abstract.search` filter and the publication-date filter `from_publication_date:1 January 1978`.

Search strategy.

Three stream-specific search strings were applied (Table 1). Each string was executed in title, abstract, and keyword fields where supported, with date and language filters set to peer-reviewed or recognised-institutional output in English. Forward and backward citation tracing from a defined set of anchor works (Sheridan and Verplank [23], Parasuraman et al. [24], Samoli et al. [31], Theofanos et al. [32], Herrmann [37], and Weill et al. [10]) supplemented database search.

Eligibility criteria.

Inclusion required: (i) that the candidate work proposed, advanced, or substantively reviewed a classification structure relevant to one of the three streams; (ii) that the work was peer-reviewed or issued by a recognised policy or research institution; and (iii) that the work was retrievable in primary form for source verification. Exclusion criteria removed candidates that addressed adjacent fields not within the present scope: explainable AI, algorithmic fairness, AI alignment, and AI evaluation each have substantial literatures of their own and are recognised as adjacent rather than overlapping with AI classification. Stream-specific selection prioritised foundational contributions and recent multi-dimensional extensions in Stream A (a mature canonical literature), the principal contemporary European and United States AI taxonomies in Stream B, and the principal subdomains of contemporary enterprise AI framework research in Stream C. The differing reference count per stream reflects the breadth of active subdomains rather than a difference in selection rigour.

Screening and selection.

Records identified through database search were progressively reduced through the PRISMA 2020 pipeline reported in Figure 1. The upstream mechanical filtering—deduplication, automated metadata exclusions (paratext exclusion, abstract present, peer-reviewed publication type, two-tier citation threshold), title-and-abstract LLM screening (Google Gemma 3 27B operating against the eligibility criteria of this section), and full-text LLM eligibility assessment via Unpaywall-retrieved open-access full text—was conducted programmatically through the screening pipeline, reducing the 24,223 OpenAlex-identified records to 314 records meeting full eligibility. The anchor-selection stage, in which 28 anchor works were selected from the 314 eligible records for narrative synthesis in Section 2, was conducted by the two authors independently; disagreements on anchor selection were resolved by discussion. The remaining 286 eligible records are retained as candidate material for the synthesis expansion in the follow-up study.

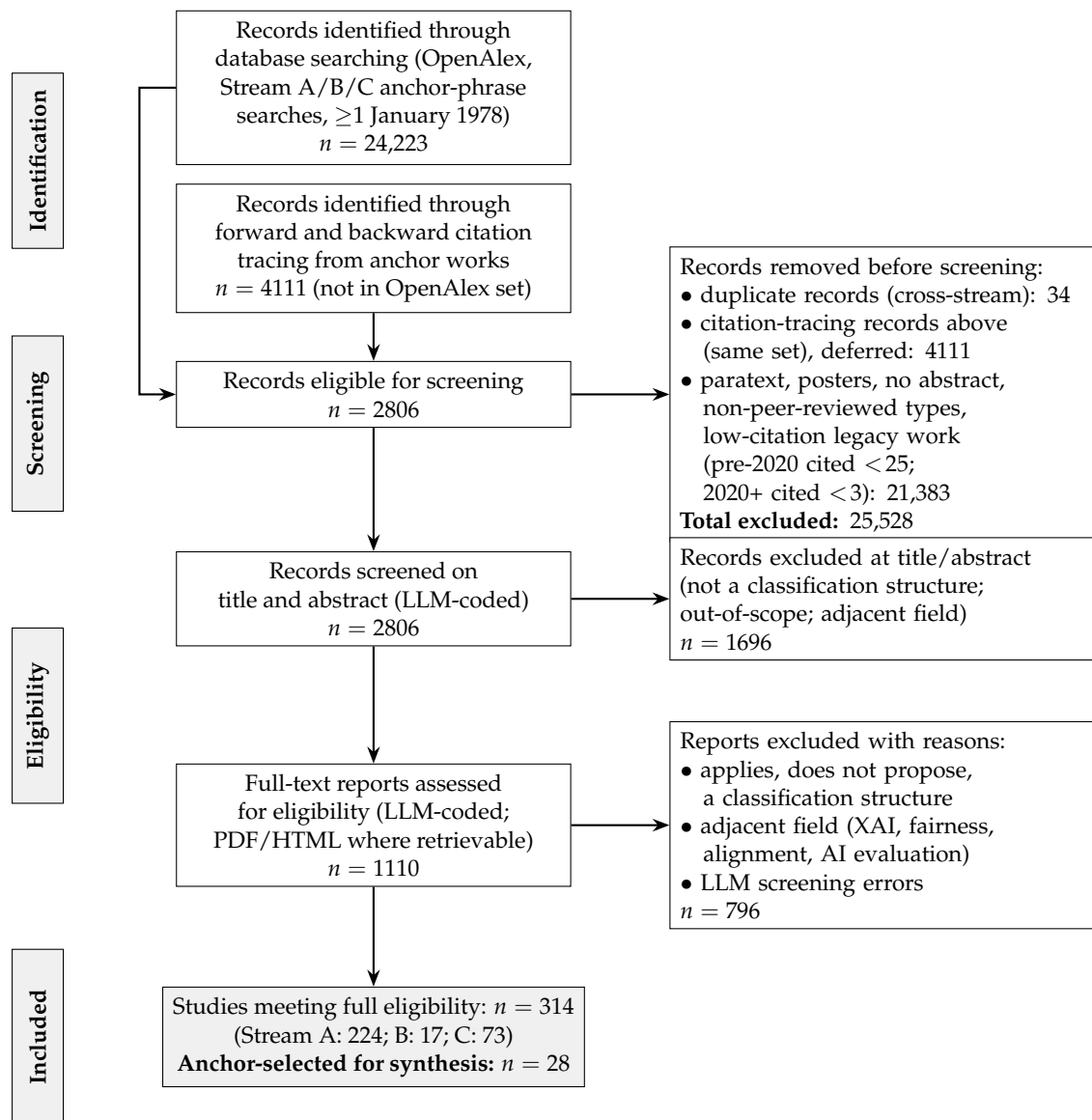


Figure 1. PRISMA 2020 flow diagram for the structured literature review. *Identification* counts were obtained on 1 May 2026 via OpenAlex using the stream-specific anchor phrases in Table 1; per-publisher cross-check counts are reported in Table 2. *Records removed before screening* comprise 34 cross-stream duplicate records, the same 4111 citation-tracing candidates (deferred as a block to the follow-up study), and automated PRISMA-2020 metadata filters (paratext exclusion, abstract present, peer-reviewed publication type, two-tier citation threshold: ≥ 25 citations pre-2020, ≥ 3 citations 2020+), yielding 2806 records eligible for screening. Title-and-abstract screening (Google Gemma 3 27B, eligibility criteria of Section 3.1) passed 1110 records. Full-text eligibility assessment retrieved open-access full text via Unpaywall (PDF 215, HTML 48, abstract-only fallback 742, none 105) and re-applied the stricter “proposes versus applies” rule via Gemma 3 27B; 314 records met full eligibility. Of these, 28 were anchor-selected for narrative synthesis in Section 2; the additional 286 eligible records are retained as candidate material for the synthesis expansion in the follow-up study. Citation tracing yielded 4111 additional candidate records aggregated across the six anchor works (4001 via OpenAlex `cites`: and `referenced_works` fields for Sheridan and Verplank [23], Parasuraman et al. [24], Theofanos et al. [32], and Herrmann [37]; 95 via Semantic Scholar for Samoili et al. [31]; 15 via manual Google Scholar and PDF endnote lookup for Weill et al. [10] MIT CISR, which is institutional grey literature outside both citation graphs). These 4111 candidates were not run through screening in the present pilot scope; concentration in domain-specific applications of anchor concepts and practitioner commentary suggests an eligibility-pass rate well below the 39.6% observed on the OpenAlex phrase-matched corpus, and the included-set is not expected to grow materially under full citation-tracing screening.

Quality appraisal.

Off-the-shelf appraisal instruments do not fit a conceptual review. AMSTAR 2 [54] and ROBIS [55] were built to appraise reviews of randomised or non-randomised healthcare-intervention studies, and they assume an effect-estimate study design that a classification review does not have; we considered them and set them aside, in line with their own stated scope. In their place, an appraisal protocol adapted to a methodological and conceptual review was applied along four criteria: (i) reporting transparency, governed by adherence to PRISMA 2020 [53]; (ii) source standing, requiring peer review or issuance by a recognised standards or policy institution; (iii) retrievability of the work in primary form for direct source verification; and (iv) contemporary relevance, enforced through the two-tier citation-and-recency threshold reported in Figure 1 (≥ 25 citations for pre-2020 works, ≥ 3 for 2020-and-later works), which functions as a proxy impact-and-currency gate on a fast-moving topic. Forward and backward citation tracing from the anchor works—the procedure automated by the Citationchaser tool [56]—was used to surface high-relevance records that string search alone would miss. A small number of older works are retained deliberately as canonical anchors whose validity is foundational rather than time-bounded (most notably Sheridan and Verplank [23] for the levels-of-autonomy lineage and Parasuraman et al. [24] for the types-and-levels integration); for the contemporary evidence base, recency was enforced by the two-tier threshold and by the addition of recent peer-reviewed journal sources during revision.

Table 2. Per-publisher identification counts for the structured literature review, derived from OpenAlex publisher-filtered queries on 1 May 2026. Each cell is the count of works whose title or abstract matches the stream’s pipe-OR phrase set and whose primary location is published by the named publisher. The IEEE, ACM, and Elsevier rows substitute for direct queries against IEEE Xplore, ACM Digital Library, and Elsevier ScienceDirect, respectively, where direct API access is unavailable. The six commercial-publisher rows together account for 10.9% of the 24,223 records identified across the three stream-specific OpenAlex queries; the remaining 89.1% are with open-access publishers (MDPI, PeerJ, JMIR), arXiv preprints, conference proceedings outside the six publishers, and institutional repositories (NIST, JRC, EIT, MIT CISR).

Publisher (OpenAlex Proxy for)	Stream A	Stream B	Stream C	Combined	% of All
IEEE (IEEE Xplore)	297	1	78	376	1.6%
ACM (ACM Digital Library)	45	0	11	56	0.2%
Elsevier BV (ScienceDirect)	1361	5	174	1540	6.4%
Springer Nature	216	1	37	254	1.0%
Wiley	386	1	33	420	1.7%
MIT Press	3	0	2	5	0.0%
Six commercial publishers, sum	2308	8	335	2651	10.9%
All venues (OpenAlex)	15,805	75	8343	24,223	100.0%

Synthesis approach.

Included works were grouped into three research streams identified *a priori* from the anchor works: (A) automation and autonomy foundations, (B) AI taxonomies and capability classifications, and (C) enterprise and generative AI frameworks. Each stream was synthesised narratively in Section 2, with each work characterised against four cross-stream descriptors: unit of analysis, axis of classification, intended audience, and presence/absence of integration with the other streams’ classifications. The cross-stream gap statement in Section 4.2 aggregates these descriptors.

3.2. Framework Construction Approach

The Enterprise AI Classification Framework was constructed as a Design Science Research (DSR) artefact following the framework of Hevner et al. [57] and the process model of Peffers et al. [58]. The classification structure is the artefact, the structural gap identified in Section 4.2 is the problem, and the design objectives derived in Section 3.2.2 below specify what a successful artefact must achieve. The taxonomy itself was developed using the iterative method of Nickerson, Varshney, and Muntermann [59] (hereafter “the Nickerson method”), which is the recognised reference method for taxonomy development in information systems research and provides explicit ending conditions against which a taxonomy can be evaluated.

3.2.1. Methodological Lineage

The DSR framing applied here extends the integrative approach validated in the authors’ prior peer-reviewed work on integrated adoption models for foundational technologies [60]. The 2022 work integrated the Technology Acceptance Model of Davis with the Capability Maturity Model (Carnegie Mellon Software Engineering Institute) in a single architecture for managing technology adoption, extending TAM with the parameters Level of Knowledge and Perceived Risk and extending CMM from five to seven levels by adding Awareness and Knowledge as the first two stages. The integrative principle of pairing two previously separate dimensions in a single architecture addressed a structural gap in the foundational-technology adoption literature in the same way that the framework proposed in the present paper addresses a structural gap in the enterprise AI classification literature. The same review-then-empirical research architecture (structured review and proposed integrated framework, followed by an empirical study across 20 industries) is applied here.

3.2.2. Design Objectives (DSR Step: Define Objectives of a Solution)

Three design objectives governed the construction of the Enterprise AI Classification Framework, derived from the gap statement in Section 4.2 and the practical risks identified in Section 1.1.

- DO1—Integration.** The framework must integrate the two dimensions that the literature addresses separately but not jointly: AI technology types and operational autonomy levels.
- DO2—Deployment-level granularity.** Each axis must be defined at the granularity at which enterprises actually identify, procure, deploy, and manage AI, rather than at the granularity of system-internal cognitive processing (as in Parasuraman et al. [24]) or capability-development trajectories (as in Morris et al. [26]).
- DO3—Shared-taxonomy tractability.** The classification must be tractable as a shared taxonomy across the heterogeneous stakeholders responsible for AI transformation, including boards, functional leaders, enterprise architects, technical teams, and risk and compliance officers.

3.2.3. Taxonomy Development Using the Nickerson Method

The Nickerson method [59] prescribes an iterative procedure with two complementary approaches per iteration—empirical-to-conceptual (E2C, examining a subset of objects and extracting characteristics) and conceptual-to-empirical (C2E, deriving characteristics from theory and applying them to objects)—terminated when the resulting taxonomy meets a set of objective and subjective ending conditions. Following Nickerson, the meta-characteristic of the present taxonomy was specified as *the deployment-level identity of an enterprise AI system*, expressed jointly in (a) what the technology does and (b) the division

of authority between human and AI in operating it. Both axes derive directly from this meta-characteristic. The development proceeded through four iterations:

Iteration 1 (C2E, types axis). Capability taxonomies from Stream B (EIT Cross-KIC [30], JRC AI Watch [31], NIST AI 200-1 [32]) were synthesised and reduced toward categories at the granularity at which enterprises procure AI as a discrete budgetary unit. Initial set: nine candidate types.

Iteration 2 (E2C, types axis). Candidate types were applied to a working set of contemporary enterprise AI deployments observed across industry. Three candidate types collapsed (audition into conversational; recommendation into predictive; robotic automation absorbed into physical AI as a deployment context), yielding the six retained types T1–T6.

Iteration 3 (C2E, levels axis). The levels-of-autonomy tradition of Stream A (Sheridan and Verplank [23], Parasuraman et al. [24], Morris et al. [26], Feng et al. [27], Porter et al. [28]) and the SAE J3016 driving-automation scale [29] were synthesised and re-projected from the individual-system unit of analysis onto the enterprise deployment unit of analysis (DO2). The synthesis yielded six levels L1–L5 with an additional L6 representing multi-system orchestration; the L5–L6 transition is non-ordinal, a property explicitly acknowledged in Section 4.3.2.

Iteration 4 (E2C, full matrix). The 6×6 matrix was applied during construction to enterprise AI deployments across supply chain, marketing, customer service, manufacturing, and finance. The application demonstrated that the matrix accommodates concurrent occupation of multiple cells by a single enterprise function. No deployment in the working set required a new type or level; the taxonomy met the conciseness and comprehensiveness ending conditions for the working set examined. The populated function-level matrices produced during construction are reserved for the empirical follow-up study described in Section 7.

The taxonomy is evaluated against the Nickerson ending conditions in Section 3.2.4.

3.2.4. Evaluation Against Nickerson's Ending Conditions

Objectiveness. Each axis is defined by a single dimension (technology function for the types axis; division of authority for the levels axis), and each cell is uniquely located by its (type, level) coordinates.

Conciseness. The framework consists of two axes of six categories each, falling within Miller's 7 ± 2 heuristic that Nickerson recommends as a conciseness anchor. The 36-cell matrix is large enough to discriminate but small enough to be held in working memory by a stakeholder.

Robustness. Different deployments of the same technology occupy different cells (e.g., a predictive AI system at L2 in credit approval vs. L4 in dynamic pricing), evidencing the discriminative capacity of the matrix.

Comprehensiveness. The four-iteration application across six enterprise functions did not surface a deployment that required a new type or level; the working set examined is bounded, and a wider corpus is examined in the computational pilot study (Section 5).

Extendibility. The matrix admits sub-typing within types (e.g., predictive AI $\rightarrow$ classification, regression, recommendation) and sub-levelling within levels (e.g., L4 $\rightarrow$ parameter ranges by risk tier) without disturbing the principal axes.

Explanatoriness. Each cell carries a statement of the division of function between human and AI, providing a directly readable description of the deployment regime to a non-technical reader.

The taxonomy met all six ending conditions for the working set examined; comprehensiveness and robustness are tested at scale in the computational pilot study reported in Section 5.

4. Results

This section presents the principal findings of the research, structured around the three research questions and the structural comparison with prior work. Section 4.1 consolidates the literature reviewed in Section 2 and answers RQ1. Section 4.2 articulates the gap that motivates the framework and answers RQ2. Section 4.3 presents the Enterprise AI Classification Framework, the principal contribution of this paper, and answers RQ3. Section 4.4 positions the framework relative to the prior frameworks reviewed in Section 2 through a structured comparison along four dimensions.

4.1. Analysis of the Literature Reviewed

The literature review presented in Section 2 covered the twenty-eight anchor works selected for narrative synthesis from the 314 records meeting full eligibility (Figure 1), spanning three research streams. Stream A, the automation and autonomy foundations literature, originates in 1978 with Sheridan and Verplank's ten-level scale of automation, develops through the two-dimensional types-and-levels model of Parasuraman, Sheridan, and Wickens, includes the configurational hybrid intelligence framework of Dellermann et al. [25], extends through the SAE J3016 driving-automation standard, the modern AGI taxonomy of Morris et al. [26], and the user-role-based agent autonomy framework of Feng, McDonald, and Zhang, and culminates in the eight-dimension INSYTE classification of Porter et al. [28]. Across the stream, the unit of analysis is the individual AI system; the stream supplies an analytical framework for characterising the autonomy and structural properties of an AI system but does not extend to the enterprise. Stream B, the AI taxonomies and capability classifications literature, comprises the principal contemporary AI capability and use taxonomies developed by European and United States policy research institutions—the EIT Cross-KIC ecosystem taxonomy with its eight capability categories, the JRC AI Watch operational definition with five core domains and three transversal domains, and the NIST AI 200-1 taxonomy of sixteen human–AI activities—supplemented by the management-oriented cognitive-technologies taxonomy of Davenport and Ronanki and the Suitability for Machine Learning rubric of Brynjolfsson and Mitchell. The stream provides structured answers to what AI is and what it does, but does not pair categories with autonomy levels. Stream C, the enterprise and generative AI frameworks literature, classifies AI at the level of the enterprise across structural taxonomies [37,38], generative AI complexity [39], agentic architectures [40–42], maturity models [10,11], enterprise operating-model accounts [44], automation typology from labour economics [45], AI management standards [46,47], and practitioner-research frameworks [48–51]. The stream addresses the enterprise as a deployer of AI but typically along a single axis; it does not integrate AI technology types with autonomy levels and does not map concrete applications simultaneously across both dimensions. RQ1 is answered: a substantial body of literature exists on AI classification, but it is distributed across three streams, each with its own unit of analysis, audience, and structural conventions.

4.2. Gap Analysis Across the Three Research Streams

Each of the three streams answers a different question. Stream A asks how autonomous a given system is. Stream B asks what type of AI it represents. Stream C asks at what stage of adoption maturity the enterprise sits. None of the three answers the question a board commissioning AI transformation actually asks: which AI technology types, at which

autonomy levels, with what concrete applications across the enterprise functional areas where AI is deployed. The gap is structural rather than incidental. The levels-of-autonomy literature characterises system autonomy in depth but does not address AI technology types or enterprise applications. The AI capability taxonomies catalogue what AI is and what it does, but pair these categories neither with autonomy levels nor with concrete enterprise applications. The enterprise and generative AI frameworks address the enterprise as a deployer but typically along a single axis: structural taxonomies of deployment categories, agentic architectures of multi-agent systems, generative AI complexity ladders, or single-axis maturity progressions. No reviewed framework maps concrete business applications simultaneously across AI technology types and autonomy levels in a form designed for enterprise transformation decision-makers.

The absence of an integrated framework is not a purely academic concern; it shapes the risk profile and the value trajectory of enterprise AI. On the risk side, the business consequences identified in Section 1.1 persist because decision-makers lack a classification architecture that answers the integrated question: investment is allocated without systematic matching of AI technology types to business applications, deployment fragments across isolated pilots, boards and technical teams operate with different vocabularies, regulatory categorisation becomes difficult to demonstrate, and enterprise AI posture becomes difficult to benchmark. On the opportunity side, an integrated classification framework would enable systematic value creation through matched AI-to-application deployment, structured innovation across the enterprise rather than within isolated functions, and transformation programmes that connect board-level intent to business application reality. The gap in the literature is therefore both a risk gap and an opportunity gap. RQ2 is answered: the structural gap across the three research streams is the absence of any framework that integrates AI technology types with autonomy levels in a single classification structure populated with concrete enterprise applications. The Enterprise AI Classification Framework, presented in Section 4.3, is proposed as the response to this gap.

4.3. The Enterprise AI Classification Framework

The framework integrates two dimensions of enterprise AI in a single structure: the types of AI technology that an enterprise deploys, and the levels of autonomy at which these technologies operate. The framework is organised around a 6×6 matrix of 36 combinations, each describing the operational regime under which a given AI technology type operates at a given autonomy level. The matrix is the framework's primary artefact, designed for direct use by business leaders commissioning and directing enterprise AI transformation. Section 4.3.1 introduces the first axis, the six AI technology types. Section 4.3.2 introduces the second axis, the six autonomy levels. Section 4.3.3 presents the full 36-combination matrix. Section 4.3.4 describes how a business decision-maker applies the framework. Section 4.3.5 addresses the location of agentic AI on the framework.

4.3.1. Six AI Types

The first axis, the types axis, classifies AI technologies by the type of function they perform at the enterprise deployment level. Six types are distinguished, each representing a distinct type of AI technology that an enterprise can identify, budget for, procure, deploy, and manage as a unit. The types are not mutually exclusive in application; a single enterprise deployment may combine multiple types, as in a customer service platform that integrates conversational AI for user interaction, predictive AI for escalation routing, and generative AI for response drafting. The types are mutually distinct as types of technology.

T1 Decision AI.

Decision AI addresses planning, scheduling, and resource optimisation. The defining characteristic is that the system determines what action to take, when, and with what resources, based on an explicit objective function. Decision AI is deployed in enterprise contexts including supply chain planning, workforce scheduling, logistics routing, and production planning, where the task is to compute an optimal or near-optimal allocation of resources against constraints.

T2 Predictive AI.

Predictive AI addresses forecasting, classification, and recommendation. The defining characteristic is that the system predicts an outcome or classifies an instance based on patterns learned from historical data. Predictive AI is deployed in enterprise contexts including demand forecasting, customer churn prediction, credit risk scoring, fraud detection, and predictive maintenance, where the task is to generate an inference about a future or unobserved state.

T3 Generative AI.

Generative AI addresses content creation, code generation, and synthesis. The defining characteristic is that the system produces novel artefacts that do not exist in its training data, including text, imagery, audio, video, code, and structured data. Generative AI is deployed in enterprise contexts including marketing content production, product design, software development, research and development, and knowledge synthesis, where the task is to generate original output at scale.

T4 Conversational AI.

Conversational AI addresses natural language dialogue and interactive exchange. The defining characteristic is that the system conducts an interactive conversation with a user or another system, maintaining context across turns. Conversational AI is deployed in enterprise contexts including customer service, internal helpdesks, virtual assistants, and voice interfaces, where the task is to resolve a request through dialogue rather than through a static query response.

T5 Visual AI.

Visual AI addresses image and video analysis, recognition, and processing. The defining characteristic is that the system interprets visual information to extract meaning or to produce a decision or action. Visual AI is deployed in enterprise contexts including quality inspection, medical imaging, security surveillance, document processing, and retail analytics, where the task is to derive insight or action from visual data.

T6 Physical AI.

Physical AI addresses robotics, autonomous vehicles, drones, and other systems that perceive and act in physical environments. The defining characteristic is that the system operates in the physical world rather than the digital world alone. Physical AI is deployed in enterprise contexts including warehouse automation, autonomous delivery, manufacturing robotics, agricultural drones, and surgical robotics, where the task requires sensing and acting in a physical environment.

A note on the abstraction level of the types axis is warranted now that the six categories have been introduced. The six types are not all distinguished by the same logical criterion: T1 decision and T2 predictive are distinguished by the function the system performs; T3 generative is distinguished by the nature of the output produced; T4 conversational is distinguished by the modality of interaction with the user; T5 visual is distinguished by the

modality of input the system interprets; and T6 physical is distinguished by the operating environment in which the system acts. The axis is therefore deployment-categorical rather than logically taxonomic in the strict sense: the categories are the units by which contemporary enterprises identify, budget for, procure, and govern AI as discrete deployments, rather than mutually exclusive logical partitions of the space of all AI systems. This design choice is deliberate and follows from design objective DO2 (Section 3.2.2): an axis defined at the granularity at which an enterprise actually deploys, manages, and procures AI is more useful for the framework's intended audience than a logically purer partition that fragments single procurement decisions across multiple categories or aggregates technically distinct deployments under a single label. The framework treats each type within a multi-type deployment as a separate locus on the matrix. The axis satisfies the Nickerson explanatory ending condition (each category is recognisable to a non-technical reader) and the conciseness ending condition (six categories falling within Miller's working-memory bound) at the cost of strict logical orthogonality across the axis. We retain this trade-off explicitly rather than concealing it.

The types axis is defined by the type of technological function, not by the cognitive processing stage within a single system as in Parasuraman et al. [24] and not by the capability developmental trajectory as in Morris et al. [26]. This distinction matters for enterprise application: a business leader commissioning AI transformation needs to identify which type of AI is being deployed, which is a question about the technology, not a question about the cognitive architecture of the system or its position on a capability progression. The six types represent the principal categories of AI technology that contemporary enterprises identify, procure, and deploy as distinct functional units; finer partitions risk fragmenting categories that enterprises treat as a single procurement, while coarser partitions obscure technologically distinct deployments that require different management.

4.3.2. Six Levels of Human–AI Authority and Coordination

The second axis, the levels axis, classifies AI deployment by the regime of human–AI authority and coordination under which the deployment operates. Six levels are distinguished. Levels L1 through L5 form an ordinal progression along a single dimension—the degree of human involvement in the operational loop—in the levels-of-autonomy tradition reviewed in Section 2.1, adapted from the individual-system unit of analysis that characterises that tradition to the enterprise deployment context in which business leaders commission and direct AI. Level L6 is qualitatively distinct: it characterises deployments in which multiple AI systems coordinate as an ecosystem, regardless of whether each constituent system is itself at L4 or L5. The L5–L6 transition is therefore non-ordinal—L6 is not “more autonomous” than L5 in the sense that L5 is more autonomous than L4, but rather introduces a coordination regime overlaid on the underlying single-system autonomy levels. We retain L6 on the same axis because, from the perspective of an enterprise leader commissioning AI, the question of *whether the deployment is a single system or an orchestrated ecosystem* is a sibling question to the autonomy-of-the-single-system question and is most usefully answered alongside it. Readers who prefer a strictly ordinal autonomy axis may treat L1–L5 as the ordinal autonomy scale and L6 as a coordination flag set on a deployment whose constituent systems already occupy positions L1–L5; the matrix in Section 4.3.3 accommodates either reading.

L1 Assistive.

AI provides information or analysis and the human performs the action. The human does the work and AI supports. Characteristic deployments include dashboards that surface trends for a human analyst, decision-support tools that compile relevant information

for a practitioner, and co-pilot interfaces that offer suggestions the human may accept or disregard.

L2 Advisory.

AI recommends options or courses of action and the human decides. The human decides and AI advises. Characteristic deployments include recommendation systems that propose ranked options for a human to select among, triage tools that suggest routing decisions, and planning assistants that generate candidate plans for human selection.

L3 Supervisory.

AI makes decisions and the human approves before execution. The human approves and AI decides. Characteristic deployments include credit approval systems that produce a decision for human sign-off, content moderation that flags items for human review, and clinical decision-support that proposes a course of action for clinician confirmation.

L4 Delegated.

AI acts independently within defined parameters and the human monitors. The human monitors and AI acts. Characteristic deployments include dynamic pricing that adjusts prices within bounded ranges, fraud detection that blocks transactions meeting defined criteria, and automated content publication within approved templates and topics.

L5 Autonomous.

AI operates without real-time human intervention. The human sets goals and AI executes. Characteristic deployments include algorithmic trading systems operating within risk envelopes, autonomous warehouse robotics fulfilling orders end-to-end, and closed-loop quality control systems that detect, decide, and act without human involvement in the operating cycle.

L6 Orchestrated.

Multiple AI systems coordinate together in an integrated workflow. The human oversees the ecosystem and the AI systems collaborate. Characteristic deployments include multi-agent logistics platforms in which routing, scheduling, and inventory systems coordinate autonomously, integrated customer experience platforms in which conversational, predictive, and generative AIs cooperate across channels, and smart-factory architectures in which vision, decision, and physical AIs coordinate production.

The levels axis describes the regime of human–AI authority and coordination for a given enterprise deployment, not the capability of the AI system considered in isolation. A given AI technology can operate at different levels depending on the enterprise context, risk tolerance, and regulatory environment: a predictive AI system may operate at L2 in credit approval (recommending a decision for human review) and at L4 in dynamic pricing (adjusting prices within defined parameters). The level is a property of the deployment, not of the technology alone. The cardinality of six levels was determined through the Nickerson iteration described in Section 3.2.3 and aligns conceptually with the six-level scales of the SAE J3016 driving-automation standard [29] and the Levels of AGI framework [26], both reoriented here to the enterprise deployment context. Five levels would collapse the L5–L6 distinction between single-system autonomy and multi-system orchestration that enterprises encounter in practice; seven or more would fragment regimes that stakeholders distinguish as a single category in shared discussion.

4.3.3. The 36-Combination Types $\times$ Levels Matrix

Combining the types axis (Section 4.3.1) and the levels axis (Section 4.3.2) produces a 6×6 matrix of 36 combinations. Each combination represents a specific pairing of AI technology type and autonomy level, stated in terms of the division of function between human and AI. A combination is located by its type (row) and level (column): for example, the cell at T2 predictive AI $\times$ L3 supervisory describes a predictive AI deployment operating at supervisory autonomy, in which the system produces a decision and a human approves before execution. A business leader identifying an enterprise AI initiative locates it on the matrix by answering two questions: what type of AI is being deployed, and at what level of human authority is it operating. Table 3 presents the matrix at the abstract level, describing the operational regime in each of the 36 cells without privileging any single enterprise function. Function-specific populated matrices—in which each cell is instantiated with concrete business applications drawn from a particular enterprise function—are the form in which the framework is applied in practice and are reserved for the empirical follow-up study described in Section 7.

Table 3. The Enterprise AI Classification Framework: 6×6 matrix of AI technology types $\times$ autonomy levels, with abstract operational regime at each cell.

	L1 Assistive	L2 Advisory	L3 Supervisory	L4 Delegated	L5 Autonomous	L6 Orchestrated
T1 Decision AI	Surfaces resource and option information for the planner	Recommends plan or allocation options for human selection	Produces plan or allocation; human approves before execution	Plans and re-plans within bounded parameters; human monitors	Runs planning and allocation autonomously	Decision AIs coordinate planning across functions or networks
T2 Predictive AI	Displays forecasts or risk patterns for human review	Recommends classifications or actions; human decides	Produces classification or scoring; human approves before action	Triggers actions within defined thresholds; human monitors	Runs forecast-and-act cycles end-to-end	Predictive AIs share inferences across networked partners
T3 Generative AI	Drafts content for human authoring	Suggests content variants; human selects	Produces content artefact; human reviews and approves	Publishes content within approved templates; human monitors	Produces and releases content end-to-end	Generative AIs coordinate content across channels
T4 Conversational AI	Suggests responses for human agents	Recommends dialogue options; human selects	Handles dialogue; human approves binding actions	Resolves interactions within defined intents; human monitors	Conducts dialogue lifecycle without intervention	Coordinates across channels and back-office systems
T5 Visual AI	Highlights regions of interest for human review	Recommends visual classifications; human selects	Produces classification decision; human confirms	Executes visual classification within thresholds	Runs continuous recognition or inspection end-to-end	Visual AIs share recognition across distributed nodes
T6 Physical AI	Augments human physical work	Suggests motion paths or actions; human executes	Executes physical task; human approves before motion	Operates within defined zones; human monitors	Operates in physical environment end-to-end	Physical AI fleets coordinate operations

The matrix accommodates instantiation across enterprise functional contexts. In supply chain, all six AI types operate concurrently at different autonomy levels: decision AI at L1–L5 for planning and routing, predictive AI at L1–L5 for demand forecasting and risk flagging, generative AI at L1–L5 for supplier communications, and so on, with cross-AI coordination falling at L6. In customer service, conversational AI operates across the autonomy range from response suggestion at L1 through fully autonomous handling at L5 and multi-channel orchestration at L6. In marketing, generative AI and predictive AI span content generation, segmentation, channel orchestration, and conversion optimisation, from drafting support at L1 through autonomous campaign execution at L5. In manufactur-

ing, visual AI operates from defect highlighting at L1 through closed-loop quality control at L5. In finance, predictive AI operates from trend display at L1 through autonomous portfolio management at L5. In operations, decision AI and predictive AI cover forecasting, scheduling, quality control, and process automation, typically from advisory support at L2 through delegated execution at L4. In research and development, generative AI and predictive AI support design exploration, simulation, and knowledge synthesis, reaching supervisory-to-autonomous operation at L3–L5 in well-bounded subdomains. These functional illustrations are signposted here as one-line examples rather than presented as populated tables; the function-by-function populated matrices belong to the empirical follow-up study (Section 7). Each function carries its own characteristic combination of AI types and autonomy levels; the matrix provides a consistent classification architecture across all functions.

The matrix integrates elements that the reviewed literature addresses separately: the autonomy dimension draws on Stream A's levels-of-autonomy tradition, adapted from the individual-system unit of analysis to the enterprise deployment context; the type dimension draws conceptually on Stream B's AI capability taxonomies, refined to six categories that an enterprise can identify and deploy; and the cross-functional illustrations introduced above draw on Stream C's enterprise-context grounding. No prior framework reviewed in Section 2 integrates these elements in this form.

4.3.4. Framework Application Logic

The framework is designed to provide a shared taxonomy through which all stakeholders responsible for enterprise AI transformation can describe, locate, and discuss AI deployment in consistent terms. Without a shared taxonomy, the parties to a transformation programme each describe AI deployment in terms specific to their role, and the conversation fragments across organisational boundaries. The framework addresses this fragmentation directly. Senior leaders responsible for strategic AI investment and direction, and the practitioners who design, execute, and govern specific transformation programmes, share the framework as a common reference. Both groups encounter the same matrix and the same illustrative applications, allowing strategic intent and operational execution to align around a common view of enterprise AI deployment, which is a precondition for coherent decision-making across the organisation. The application logic proceeds in three steps.

The first step is identification. A given enterprise AI deployment is located on the matrix by determining its type of AI technology and its current level of human–AI authority. The deployment's position is fixed by the intersection of these two coordinates. Multiple deployments within the same enterprise function are located independently and may occupy different combinations in the matrix, reflecting the enterprise's actual portfolio rather than a single point estimate.

The second step is assessment. The matrix supports comparative assessment of the enterprise's current AI portfolio against three reference points: the distribution of combinations the enterprise currently occupies, the combinations occupied by peer enterprises in the same function or industry, and the combinations where additional value or risk reduction could be realised through deployment. The classification is descriptive rather than normative; the framework does not prescribe that any specific combination is the correct destination, and decisions about target combinations reflect business strategy, regulatory environment, and risk tolerance rather than framework design.

The third step is direction. The matrix supports decisions about where to invest in new AI deployment, where to advance the autonomy level of existing deployments, where to constrain or supervise deployments more closely, and where to retire or replace AI that

is poorly matched to its enterprise context. The two-dimensional structure makes these decisions visible and discussable across boards, technical teams, and functional leaders. The framework therefore functions as a classification and decision-support architecture for enterprise AI transformation.

4.3.5. Locating Agentic AI on the Framework

Agentic AI deployments—systems that decompose tasks dynamically, invoke tools, maintain persistent memory, and coordinate as multi-agent ensembles [40–42]—are an increasingly prominent form of enterprise AI in 2026 and warrant explicit treatment. The framework accommodates agentic AI through three locating moves rather than a separate axis. First, an individual agent is located on the matrix by the type of function it performs (T1–T6) and the autonomy level at which the enterprise has authorised it to operate (L1–L5). A research agent that drafts and revises analyst briefings under human approval is a T3 generative deployment at L3 supervisory; a procurement agent that issues bounded purchase orders under monitoring is a T1 decision deployment at L4 delegated. The agentic implementation pattern (LLM-based reasoning, tool invocation, memory) does not in itself change the cell on the matrix; what fixes the cell is the function performed and the operational regime granted. Second, multi-agent ensembles in which heterogeneous agents coordinate are located at L6 orchestrated, with the constituent agents simultaneously occupying their respective (T, L1–L5) cells. A customer experience platform in which a conversational agent (T4) coordinates with a predictive routing agent (T2) and a generative response-drafting agent (T3) is recorded as occupying three constituent-agent cells ($T4 \times L_n$, $T2 \times L_n$, $T3 \times L_n$) and additionally three L6 cells ($T4 \times L6$, $T2 \times L6$, $T3 \times L6$) reflecting the ensemble-level coordination regime. Third, agents whose action repertoire spans multiple types—common in generalist LLM-based agents—are recorded across the relevant rows simultaneously rather than forced into a single row, with each row recording the autonomy level applicable to that function within the deployment. The framework therefore provides a multi-cell representation for agentic AI, preserving the descriptive resolution that a single-cell forced fit would obscure.

4.4. Comparison with Prior Frameworks

The framework is positioned in relation to the prior work reviewed in Section 2 through a structured comparison along four dimensions: whether the framework classifies AI by AI technology type, whether it classifies AI by autonomy level, whether it integrates both axes, and whether it is designed for enterprise deployment contexts. The four dimensions correspond to the central classification question of each of the three research streams reviewed in Section 2 (Stream A's autonomy axis, Stream B's AI technology types, Stream C's enterprise deployment focus) plus the cross-stream integration that motivates the present paper; the criteria are derived from the structure of the literature rather than selected post hoc to favour the proposed framework. Table 4 presents the comparison.

The comparison illustrates the structural position of the framework. No prior framework reviewed combines positive answers across all four dimensions. Stream A frameworks address autonomy levels but typically without an AI technology type classification, and Parasuraman et al. [24] is the closest historical precedent for a types $\times$ levels integration but uses cognitive processing stages within an individual system rather than AI technology types deployed across an enterprise. Stream B frameworks address AI technology classification but without an autonomy levels axis. Stream C frameworks address enterprise deployment but typically along a single axis, with neither the AI technology type breakdown nor the autonomy level dimension present in Stream A and B work. The framework is positioned as the integration of these three lines of contribution into a single

architecture designed for AI-driven enterprise transformation. The structural comparison presented here completes the answer to RQ3 begun in Section 4.3: the new AI classification framework proposed in this paper occupies a structural position in the AI classification literature that no prior framework occupies.

Table 4. Comparison of the proposed framework with prior AI classification frameworks.

Framework	AI Types	Autonomy Levels	Both Axes Integrated	Enterprise Focus
Sheridan and Verplank [23]	No	Yes (unidimensional scale)	No	No
Parasuraman et al. [24]	Partial (cognitive processing stages)	Yes	Yes (cognitive types $\times$ levels)	No
Dellermann et al. [25]	No	No (configurational)	No	Partial
Morris et al. [26]	Partial (capability dimensions)	Yes	Partial	No
Feng et al. [27]	No	Yes (user-role focused)	No	No
Porter et al. INSYTE [28]	No (functional rather than typological)	Yes (two sub-dimensions under Operational Independence: Intervention and Oversight)	No (no types $\times$ levels integration)	No
EIT Cross-KIC [30]	Yes (multi-dimensional structural taxonomy: five dimensions including AI Capabilities and Enterprise Functions)	No	No (no autonomy axis)	No
Samoili et al. JRC AI Watch [31]	Yes (operational AI domain/subdomain taxonomy)	No	No	No
Theofanos et al. NIST AI 200-1 [32]	Yes (activity categories)	No	No	No
Davenport and Ronanki [33,34]	Yes (three cognitive task categories: process automation, cognitive insight, cognitive engagement)	No	No	Partial (management-oriented)
Brynjolfsson and Mitchell [35,36]	No (task-suitability criteria, not AI categorisation)	No	No	No (labour economics unit of analysis)
Herrmann [37]	Yes (Euler diagram of fields and subfields)	No	No	Yes
Bashir [38]	No	No (single-axis SEAI hierarchy)	No	Yes
Stein [39]	Partial (genAI only)	Partial (technical autonomy as one of five perspectives)	No	Yes
Sapkota et al. [40]	Partial (agentic AI only)	Partial	No	Partial
Weill, Woerner, and Sebastian [10]	No	No (single-axis maturity)	No	Yes
Enterprise AI Classification Framework	Yes (six AI technology types)	Yes (six autonomy levels)	Yes (6 $\times$ 6 matrix)	Yes

Note. The agentic-architecture frameworks of Panigrahy [41] and Arunkumar et al. [42] are addressed in the Stream C review (Section 2.3) but are not included in Table 4 because they characterise the internal architecture of multi-agent AI systems rather than the four classification dimensions used in the table.

5. Computational Pilot Study

This section reports a computational pilot study designed to test the Enterprise AI Classification Framework against four baseline frameworks before the interview-based empirical validation described in Section 7. The pilot study is conducted in two parts. Study A is a case-based coding validation across a corpus of publicly documented enter-

prise AI deployments, measuring the framework's coverage, inter-coder reliability, and discriminative power against baselines. Study B is a multi-persona classification stress test that operates as a pre-specification of role-specific hypotheses for the planned 250-interview validation. The locked rulebook v1.0, the LLM coder prompts, the pre-specified hypotheses, and the analysis plan constitute the pre-registration package; the package was filed as a pre-registration on the Open Science Framework on 12 May 2026 under a twelve-month embargo, ahead of the full 240-case run that will produce the Study A metrics in the follow-up empirical study.

5.1. Study A: Case-Based Coding Validation

5.1.1. Design

A corpus of approximately 240 publicly documented enterprise AI deployments is assembled from six source classes: peer-reviewed and management-research case studies (MIT CISR, MIT SMR, Harvard Business Review, California Management Review), publicly published consultancy case studies (McKinsey, BCG, Bain, Deloitte, Accenture, KPMG, EY), industry analyst entries (Gartner Hype Cycle, IDC, Forrester), vendor case studies (after deduplication), regulator and standards-body filings, and news-reported deployments. Inclusion requires that the deployment is at enterprise scale, that the source contains sufficient detail to determine what the AI does and how human–AI authority is distributed, that the source is retrievable in primary form, that the publication date is 2020 or later, and that the source is available in English, German, French, or Spanish. The corpus is stratified across industry, enterprise size, and geography. The full inclusion and exclusion criteria are specified in the locked Study A coding rulebook accompanying this paper.

5.1.2. Coding Protocol

Each case is independently coded by two large language model coders (Claude Opus 4.7 and NVIDIA Nemotron 3 Super 120B-A12B) using a locked prompt that operationalises the Enterprise AI Classification Framework and four baseline frameworks—the Parasuraman types-and-levels model [24], the NIST AI 200-1 use taxonomy [32], the SEAI hierarchy [38], and the MIT CISR Building Enterprise AI Maturity model [10]. A 12% sample is independently coded by the second author blind to the LLM codes, providing a calibration check on the LLM-LLM agreement. For the 20-case pilot, three cases (P-001, P-006, P-014) were spot-checked by the second author at the 15% rate; the second author's classifications matched both LLM coders on all three cases (3/3 cell-set agreement). The full 240-case empirical run will extend this to a random sample of approximately 28–30 cases (12%), drawn after the LLM coding is complete so that the human classifier remains blind. Disagreements between the two LLM coders are resolved by a third LLM (Google Gemma 3 27B) acting as adjudicator with a separate prompt that frames the disagreement and asks for resolution given the case description; a cross-vendor adjudicator was chosen over a same-model second-pass to reduce the shared-bias risk between coder and adjudicator. The locked v1.0 pilot used Gemma 3 27B as adjudicator under design decision D2 (free-tier API quota); a conditional reversion to Gemini 3 for the adjudicator and to Gemini 2.5 Pro alongside Claude Opus 4.7 as the two coders is noted as future work for the full 240-case run if budget permits. The adjudicated codes are the final classifications for analysis.

Four design choices in the coding protocol are locked in the rulebook and recorded explicitly here. *First*, where the source does not specify human approval procedures, the coder defaults to L3 supervisory for high-risk applications (regulated industries, safety-critical, employment, credit) and to L4 delegated for low-risk applications; this reduces the unclassifiable rate while introducing a known L3/L4-band bias that is reported in the analysis. *Second*, coder confidence is recorded for the framework classification only, not for

each baseline, in the interest of operational simplicity at this stage of the pilot. *Third*, the corpus stratification targets (eight industries, three enterprise-size bands, four geographies) are validated against actual source availability through a 20-case pilot extraction before the full corpus is assembled. *Fourth*, all four design decisions are recorded in a versioned design decision log alongside the rulebook and were locked at OSF pre-registration on 12 May 2026; the original choices are not overwritten so that the audit trail remains intact.

5.1.3. Hypotheses and Metrics

Four pre-specified hypotheses test the framework's structural claims.

H1 Coverage. The proportion of corpus cases assigned a non-unclassifiable code under the framework exceeds the proportion under each of the four baselines by at least 10 percentage points.

H2 Reliability. Krippendorff's α for the framework's type and level axes is at least 0.70 and is at least as high as the corresponding α for each baseline's classification axes.

H3 Discriminative power. The Shannon entropy of the empirical case distribution across occupied cells of the framework, normalised by $\log_2$ of the number of occupied cells, exceeds the corresponding normalised entropy for each baseline.

H4 Heatmap. The empirical density of cases across the framework's 6×6 matrix is reported as a heatmap, separately by industry on the full corpus, as the principal empirical artefact of Study A. Figure 2 reports the heatmap for the $n = 20$ pilot corpus; the per-industry panels are produced from the full corpus in the empirical follow-up study.

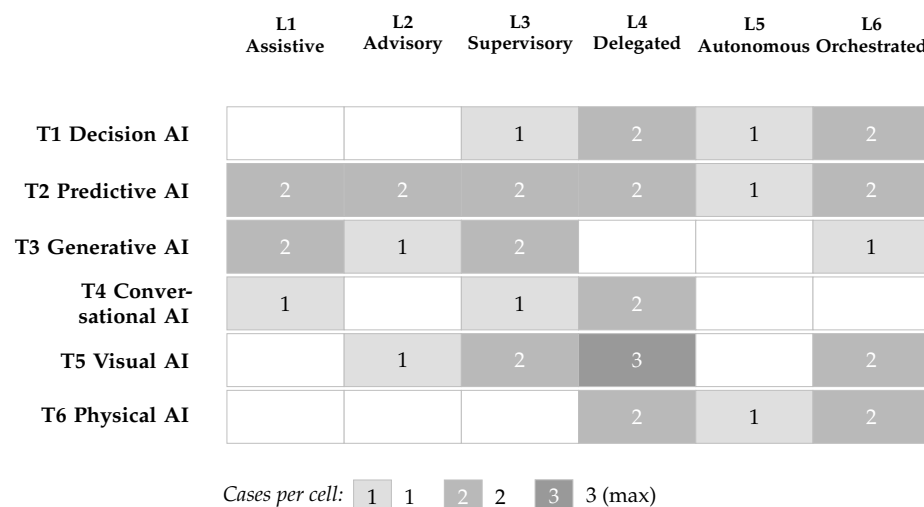


Figure 2. H4 cell-density heatmap from the Study A pilot corpus ($n = 20$ adjudicated cases). Each cell reports the number of pilot cases whose final adjudicated classification occupies that (type, level) coordinate, with shading proportional to that count (40 cell occupations across the 20 cases; 24 of 36 cells occupied). Levels L1–L5 are taken directly from the adjudicated cell codes; the L6 column is derived from the three cases coded as orchestrated, each contributing one L6 cell per constituent type under the multi-cell representation of Section 4.3.5. The pilot corpus is not stratified by industry; the per-industry panels described under H4 are produced from the full corpus in the empirical follow-up study (Section 7). At the pilot scale the modal band is the L3–L4 (supervisory–delegated) region, consistent with the L3/L4 default rule recorded in the coding protocol.

A fifth analysis (H5) examines the residual of cases coded unclassifiable under the framework as a qualitative analysis: residual cases are grouped by emergent theme to surface concrete refinements to the framework rather than fundamental gaps.

5.2. Study B: Multi-Persona Classification Stress Test

Study B operates as a pre-specification of role-specific hypotheses for the planned 250-interview validation. Twelve personas matching the planned interview population were operationalised as large language model role specifications: chief executive officer, chief information officer, chief technology officer, chief financial officer, head of supply chain, head of marketing, head of customer service, chief risk or information security officer, head of human resources, non-executive board member, AI lead, and enterprise architect. All persona classifications were produced by Google Gemini 2.5 Flash Lite, accessed through the free-tier Google AI Studio quota. Each persona classified the same 30 standardised AI deployment vignettes spanning the framework's matrix and including a deliberate subset of edge cases (the L1–L2 boundary, multi-agent orchestration, multi-modal foundation models, prohibited-tier emotion recognition, single-deployment-without-enterprise-context). For each (persona, vignette) pair, the persona returned (a) the current-state framework classification, (b) the target-state classification at three-year horizon, and (c) a brief justification.

Three pre-specified hypotheses were defined for Study B. *H6(a)* predicted that cross-persona agreement on the current-state classification would meet a Krippendorff's $\alpha \geq 0.70$ threshold, evidencing the framework's shared-taxonomy claim. *H6(b)* predicted that cross-persona agreement on the target-state classification would fall below $\alpha < 0.70$, evidencing role-specific divergence on strategic direction. *H6(c)* predicted that the risk-conservative persona pair (chief financial officer plus chief risk officer) would target autonomy levels at least one level lower than the technology-aggressive persona pair (chief technology officer plus AI lead) on finance and credit deployments, controlling for vignette.

Study B is a synthetic pre-specification check rather than an empirical validation: the twelve personas are LLM-simulated against author-specified role descriptions, and the thirty vignettes are author-written. The result characterises the internal coherence of the role-divergence hypothesis specifications and the operational behaviour of the LLM-persona instrument, not the behaviour of real role-holders, which is the subject of the planned 250-interview validation. At this synthetic scale, all three pre-specified hypotheses behaved as designed: cross-persona α on the current-state classification was 0.730, just above the 0.70 threshold; cross-persona α on the target-state classification was 0.402, well below the same threshold; and the *H6(c)* divergence on finance and credit deployments measured +1.10 levels in the predicted direction (CFO and CRO mean target level 3.40; CTO and AI lead mean target level 4.50; ten observations per group across the five finance and credit vignettes), with a bootstrapped 95% confidence interval on the difference of $[+0.70, +1.50]$ ($n_{\text{boot}} = 2000$, paired persona-group resampling) excluding zero and the *H6(c)* pre-specified ≥ 1.0 threshold contained in the CI. The role-specific direction-of-travel pattern between risk-conservative and technology-aggressive roles therefore emerges in the synthetic instrument at the magnitude pre-specified as the empirical-study target; whether real role-holders reason as the LLM personas predict is the central empirical question of the planned 250-interview validation. The full Study B protocol, vignette set, persona specifications, and per-vignette breakdown are part of the OSF pre-registration package described in the Data Availability Statement (osf.io/6jnza), filed under a twelve-month embargo and to be released on OSF upon publication.

5.3. Pilot Results

The pilot was conducted in three stages culminating in a full 20-case two-coder + adjudicator run at locked rulebook v1.0. A 20-case single-coder rulebook stress test (Coder A only) was run first against rulebook v0.3 and surfaced five operational refinements (LLM dual-use coding, T2/T5 boundary tie-breaker, L6 threshold tightening, Defense-industry

stratification, low-confidence reason capture). A 20-case two-coder + adjudicator pilot at v0.3 produced α on the framework's primary-cell-type axis of 0.578 and on the level axis of 0.729; coverage was 100% framework on both coders versus 75% on the SEAI and MIT CISR baselines on Coder A's careful application of the unclassifiable rule. Five further findings from the v0.3 two-coder pilot were folded into rulebook v0.4 (LLM dual-use rule strengthened with worked examples; T2/T5 tie-breaker added; L6 negative markers added; adjudicator output schema constrained; SEAI/MIT-CISR unclassifiable instruction strengthened). The full 20-case v0.4 (later locked as v1.0 for the OSF pre-registration package) produced the headline metrics in Table 5.

Table 5. Study A pilot metrics at locked rulebook v1.0 (full 20 cases, Coder A: Claude Opus 4.7; Coder B: NVIDIA Nemotron 3 Super 120B-A12B; Adjudicator: Google Gemma 3 27B). Bootstrapped 95% confidence intervals computed by case-level resampling with replacement, $n_{\text{boot}} = 2000$.

Metric	Value	95% CI	Target	Verdict
α primary cell type, nominal	0.809	[0.582, 1.000]	≥ 0.70	H2 met (point)
α primary cell level, ordinal	0.802	[0.530, 0.935]	≥ 0.70	H2 met (point)
α primary cell T + L pair, nominal	0.628	[0.395, 0.832]	≥ 0.70	close, below
Cell-set agreement (full equality)	12/20 (60%)	—	—	
Framework coverage, Coder A	100%	—	—	
Framework coverage, Coder B	100%	—	—	
SEAI baseline coverage, Coder A	70%	—	—	
SEAI baseline coverage, Coder B	0%	—	strict v1.0 rule applied	
MIT CISR baseline coverage, Coder A	70%	—	—	
MIT CISR baseline coverage, Coder B	0%	—	strict v1.0 rule applied	
Multi-cell occupation rate, both coders	40%	—	matches design	
L6 attribution rate, Coder A/Coder B	15%/10%	10–15%	in design band	
Disagreements adjudicated	8 of 20	—	—	

The metrics below are point-estimate signals at the locked v1.0 rulebook on $n = 20$ cases; the bootstrapped 95% confidence intervals are wide and include values below the pre-specified thresholds. The full 240-case run in the follow-up study is expected to tighten the intervals substantially. H1 (coverage) is supported at the pilot scale: framework coverage at 100% on both coders exceeds the average baseline coverage of approximately 35% by approximately 65 percentage points, well above the pre-specified 10-percentage-point threshold. The Coder B 0% baseline coverage on SEAI and MIT CISR reflects the strict v1.0 unclassifiable instruction operating on cases lacking the enterprise-strategic context that those baselines require; reported per-coder (Coder A SEAI 70%, Coder A MIT CISR 70%), the framework-versus-baseline coverage gap remains substantial but is better calibrated than the cross-coder average suggests. H2 (reliability) is met at the point estimate on both axes (type $\alpha = 0.809$, level $\alpha = 0.802$). The bootstrapped 95% confidence intervals on the type and level axes ([0.582, 1.000] and [0.530, 0.935]) extend below the 0.70 pre-specified threshold; the lower bounds reflect the small sample size ($n = 20$) of the pilot, and the full 240-case run in the follow-up study is expected to tighten the CIs substantially. The combined T + L pair α of 0.628 (CI [0.395, 0.832]) falls short of the 0.70 threshold at the point estimate; analysis of the eight disagreements shows that level-axis disagreements cluster predominantly on the L1 (assistive) versus L2 (advisory) boundary, the residual rulebook ambiguity that will be the principal site of refinement in the full-corpus run. The blind human spot-check on three of the twenty cases (15%) returned 3/3 cell-set

agreement with both LLM coders, consistent with the framework's reliability claim on the easy-case subset of the corpus; the full random-sample human spot-check is deferred to the empirical study. Type-axis disagreements (four of eight) are all on multi-cell edge cases where the framework's multi-cell representation accommodates either coder's enumeration of subsystems; none reflect a structural failure of the type axis. The Coder B SEAI and MIT CISR coverage of 0% reflects the strict v1.0 unclassifiable instruction operating as designed: where the case description is a single deployment without explicit enterprise-strategic context, Coder B abstains rather than force-fits.

5.4. Limitations of the Synthetic Evaluation

Four limitations of the synthetic evaluation should be acknowledged. First, large language model coders and personas are not substitutes for human experts. The two-LLM coding pair plus human spot-check protocol is designed to detect divergence between LLM consensus and human judgment, but does not eliminate the possibility that both LLM coders share systematic biases the spot-check fails to catch. Second, the corpus is restricted to publicly documented deployments, which over-represents deployments that organisations choose to publicise and under-represents quietly operating production systems and failed deployments. The empirical validation in the follow-up paper addresses this exposure by sampling deployments through the 250 interview partners rather than through the public-source corpus. Third, Study B's persona simulations are grounded in role specifications constructed by the authors and may reflect the authors' assumptions about role-specific reasoning patterns; the 250-interview study will test whether real role-holders reason as the LLM personas predict. Fourth, the L1 versus L2 boundary on the autonomy axis was identified in the v0.4 pilot as the principal residual ambiguity in the rulebook; this boundary is operationally subtle (the difference between an AI that informs a decision the human would have made anyway and an AI that proposes options among which the human selects) and is the most likely site of inter-coder disagreement at the full corpus run. The synthetic evaluation is therefore positioned as a computational pilot study and a pre-specification mechanism, not as a substitute for the empirical validation.

6. Discussion

6.1. Theoretical and Practical Contributions

The theoretical contribution is structural: AI technology types and autonomy levels are integrated into a single classification structure populated with concrete enterprise applications, an architecture none of the three reviewed streams provides. Stream A classifies the autonomy of automated systems on an ordinal scale at the unit of the individual system. Stream B classifies AI as a class of technology along categorical taxonomies of capability or activity. Stream C classifies enterprise AI either as a staged adoption progression or as a structural taxonomy of deployment categories. Each dimension has matured within its own stream, but none combines the three. The framework integrates the autonomy, technology-type, and enterprise-context dimensions in one architecture: the autonomy question is moved from the individual-system unit of analysis to the enterprise deployment context, the technology-type classification is refined into six categories an enterprise can identify and deploy, and the result is grounded in concrete business application examples across enterprise functional contexts.

The practical contribution is an operational framework for enterprises commissioning AI transformation. The framework provides a shared taxonomy through which boards, functional leaders, and technical teams can describe, locate, and discuss AI deployment in consistent terms, addressing the misalignment between these stakeholder groups identified in Section 1.1 as a structural risk of contemporary enterprise AI. The framework supports

portfolio assessment, peer benchmarking, and the systematic identification of investment, advancement, constraint, and retirement decisions across the enterprise AI deployment landscape. The framework is a classification artefact rather than a governance artefact; integrating AI classification with enterprise IT governance practice is a distinct undertaking left to future work, and recent systematic reviews map that governance agenda and its open questions [21,22].

6.2. Practical Implications and Intended Beneficiaries

A framework earns its keep only when someone uses it. It is worth saying plainly, then, who uses this one and what they get out of it. Seven groups benefit, each in a different way.

Boards and senior executives gain a portfolio-level instrument for AI investment governance. By locating every enterprise AI initiative on the 6×6 matrix, a board can see at a glance where its AI spend is concentrated, whether the portfolio is skewed toward low-autonomy assistive deployments or toward higher-risk autonomous ones, and where investment is missing. The practical value is a defensible basis for capital-allocation and risk-appetite decisions that the value-gap evidence [12,13] shows are otherwise made without a structured view.

Functional and business-unit leaders gain a planning instrument for their own function. By instantiating the matrix for, say, supply chain or customer service, a functional leader can plan the progression of a deployment from advisory toward delegated or autonomous operation as confidence and controls mature, and can benchmark that progression against peers.

Enterprise architects gain a mapping notation. The matrix provides a consistent vocabulary for documenting the AI estate, identifying redundancy and integration opportunities, and recording the multi-cell footprint of agentic and multi-agent systems (Section 4.3.5).

Risk, compliance, and security officers gain a categorisation aid for regulatory exposure. The type-and-level coordinates map naturally onto the risk-tiering logic of the EU AI Act and comparable regimes, supporting the categorised, auditable deployment inventory those regimes require; the framework interoperates with, rather than replaces, the governance instruments reviewed in Section 1.1 and [21,22].

AI vendors and consultancies gain a positioning vocabulary, able to describe an offering by the cells of the matrix it occupies and the autonomy transitions it enables, rather than by marketing category.

Policymakers and standards bodies gain a common reference structure that bridges the system-level autonomy standards of Stream A and the capability taxonomies of Stream B at the enterprise unit of analysis where deployment, and therefore regulation, actually bites.

Researchers gain an empirical instrument: the matrix is the coding scheme for the computational pilot study (Section 5) and the planned 250-interview validation, and is reusable as a measurement frame for future enterprise AI studies.

6.3. International Comparison of Findings

The Enterprise AI Classification Framework is derived from AI technology type and autonomy level, neither of which is jurisdiction-specific. Its applicability is therefore global. What varies by jurisdiction is the distribution of enterprise deployments across the matrix, the regulatory environment in which they operate, and the policy responses to the value gap. The three jurisdictions discussed below illustrate the latter; the planned 250-interview validation will characterise the former across twenty industries and multiple regions.

The structural gap is global; the appetite for closing it is not. The evidence shows sharp regional differences in how hard enterprises are actually pushing on AI, and those differences shape how the framework would be applied in each setting.

The strategic-investment signals reviewed in the introduction are cross-national to begin with: China's 15th Five-Year Plan, the European Union's InvestAI and AI Continent programmes, and the national commitments announced at the Paris AI Action Summit together show governments on three continents treating AI classification and deployment as industrial-policy priorities. The value gap shows up across jurisdictions, not within one: the global surveys behind the motivating statistics span North America, Europe, and Asia, and the peer-reviewed evidence is distributed the same way, with productivity studies run in the United States [14,15] and business-value and capability work drawing on European and international samples [12,13]. Adoption *intensity*, however, diverges sharply. Official European statistics for 2024 [61] record wide cross-country dispersion in enterprise AI use—from above one quarter of enterprises in the Nordic and Benelux economies to low single digits in parts of central and eastern Europe—and firm-level studies attribute this dispersion to differences in digital capability, firm size, and external environment rather than to the technology itself. The study of European small and medium-sized enterprises by Arroyabe et al. [62], drawing on a sample of more than twelve thousand firms, finds digital and innovation capabilities to be the dominant determinants of adoption, and the European process-industry maturity assessment of Fornasiero et al. [43] finds maturity to vary by operational dimension even within a single sector and region.

Two implications follow for the framework. First, because the structural gap—the absence of an integrated type-and-level classification—is present wherever enterprises deploy AI, the framework's relevance is not confined to any one regulatory or economic setting; the same 6×6 structure applies whether an enterprise operates under the EU AI Act, under the more sectoral United States approach, or under China's state-directed programme. Second, because adoption intensity and maturity differ across regions, the *distribution* of an enterprise's deployments across the matrix is expected to differ systematically by jurisdiction: enterprises in high-maturity economies are likely to occupy higher-autonomy and orchestrated cells more densely, while enterprises in lower-maturity economies are likely to cluster in the assistive and advisory columns. The planned 250-interview validation across 20 industries is designed to sample internationally so that this cross-jurisdictional variation in the matrix distribution can be measured directly rather than assumed, which is the natural next step in establishing the framework's external validity.

6.4. Limitations

Several limitations should be acknowledged. First, the structural integration of AI technology types with autonomy levels presented in this paper is grounded in the comparative literature analysis of Section 2 and the construction approach described in Section 3; this paper's empirical content is limited to the computational pilot study reported in Section 5, with interview-based validation deferred to the follow-up paper described in Section 7. Second, the framework's full applicability is realised through detailed application matrices for each enterprise function with concrete combination content tailored to the specific deployment patterns of that function. The present paper presents the framework as an abstract 6×6 matrix (Table 3) with cross-functional signposting; the development of populated function-level matrices for the full set of enterprise functions (supply chain, marketing, customer service, manufacturing, finance, human resources, research and development, information technology operations) is reserved for the empirical follow-up study described in Section 7. Third, the cross-functional examples illustrating how the matrix accommodates instantiation may evolve as AI technology matures and as new enterprise use cases

emerge, although the matrix architecture is expected to remain stable; the framework's illustrative content is therefore time-bounded in a way that the framework structure is not.

7. Conclusions

Investment in AI-driven enterprise transformation has now reached a scale that demands a structured classification framework. Enterprises across every sector deploy AI across the six AI technology types examined in this paper, at autonomy levels ranging from assistive support to fully autonomous operation; the pace of investment has outrun the conceptual frameworks needed to navigate it coherently.

This paper has addressed three research questions. RQ1 asked which AI classification frameworks and approaches are available for AI-driven enterprise transformation. Section 2 synthesised twenty-eight anchor works selected from 314 records meeting full eligibility across the automation and autonomy literature, the AI taxonomies and capability classifications literature, and the enterprise and generative AI frameworks literature; Section 4.1 consolidated the answer. RQ2 asked what current research gaps exist in AI classification frameworks for AI-driven enterprise transformation. Section 4.2 identified the structural gap as the absence of any framework that integrates AI technology types with autonomy levels in a single architecture populated with concrete enterprise applications. RQ3 asked which new AI classification framework can address the identified gaps. The Enterprise AI Classification Framework, proposed in Section 4.3 and defended through structured comparison in Section 4.4, is the answer offered by this paper. The framework combines six AI technology types with six autonomy levels in a 6×6 matrix of thirty-six combinations, each describing the operational regime under which a given AI technology type operates at a given autonomy level, providing a shared taxonomy through which all stakeholders responsible for AI transformation can describe, locate, and discuss AI deployment in consistent terms.

The contribution of the framework is theoretical, methodological, and practical. Theoretically, it integrates three previously separate research conversations into a single classification architecture. Methodologically, it translates a technically grounded classification into a form designed for business use, populated with applications business leaders recognise. Practically, it provides an operational framework for enterprises commissioning AI transformation, supporting portfolio assessment, peer benchmarking, and systematic decisions on investment, advancement, constraint, and retirement of AI deployments. The framework has been operationalised into an enterprise platform that allows organisations to analyse, design, and develop their AI portfolio across the 6×6 matrix; the platform is the first working implementation of the classification and demonstrates its applicability to real-world enterprise AI decision-making.

The forward research plan proceeds in three concrete steps. The first step is empirical validation of the framework with decision-makers, enterprise architects, transformation practitioners, and risk and compliance officers across 20 industries, assessing whether the structural assumptions of the framework hold in practice and whether the framework supports the shared-taxonomy purpose for which it is designed. The validation will apply the integrated TAM $\times$ CMM methodology established in the authors' prior peer-reviewed research on integrated adoption models for foundational technologies [60], in which a structured literature review and proposed integrated framework were validated in a subsequent empirical study across 20 industries and 125 business leaders. That prior delivery, conducted by the authors within the same review-then-empirical two-paper architecture applied here, supports the feasibility of the 250-interview target as a calibrated extension of the delivered 125-leader baseline rather than a speculative scale-up. The second step is the development of detailed function-level matrices for the full set of enterprise

functions—supply chain, marketing, customer service, manufacturing, finance, human resources, research and development, and information technology operations—in which each cell of the framework matrix (Table 3) is instantiated with concrete business applications drawn from the function’s deployment patterns. The third step is longitudinal application of the framework as enterprise AI deployments evolve, tracking how organisations migrate across combinations in the matrix over time and documenting the patterns of advancement and retirement that emerge.

Author Contributions: Conceptualization, N.B. and V.S.; methodology, N.B. (structured literature review, framework construction) and V.S. (Design Science Research framing, taxonomy-development methodology, synthetic-evaluation design); software, V.S.; validation, N.B. (human spot-check on framework coding rules) and V.S. (computational pilot study metrics, reliability analysis); formal analysis, V.S.; investigation, N.B. (literature review across the three research streams; framework axes and matrix construction) and V.S. (computational pilot Studies A and B; full-corpus screening pipeline); data curation, V.S.; writing—original draft preparation, N.B. (Section 2 Structured Literature Review, Section 4.3 Enterprise AI Classification Framework, Section 4.4 Comparison with Prior Frameworks), V.S. (Section 3 Methodology, Section 5 Computational Pilot Study), and N.B. and V.S. jointly (Section 1 Introduction, Sections 4.1 and 4.2 results consolidation and Gap Analysis across the Three Research Streams, Section 6 Discussion, Section 7 Conclusions); writing—review and editing, N.B. and V.S.; supervision, V.S.; project administration, V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in the conceptual portion of this study. The pre-registration package for the computational pilot study reported in Section 5—the locked coding rulebook, the LLM coder prompts, the pre-specified hypotheses, and the analysis plan—was filed on the Open Science Framework (OSF) on 12 May 2026 as registration `osf.io/6jnza` under a twelve-month embargo that lifts on 12 May 2027; per OSF policy, a DOI for the registration will be minted upon embargo expiry. OSF can confirm filing date and contents to the editorial office on request. The corpus and coded dataset will be released on OSF upon publication.

Acknowledgments: The authors thank the open-research infrastructure that made the computational pilot study feasible at zero cost: the OpenAlex API (<https://api.openalex.org>, accessed on 25 June 2026; OurResearch, Inc., Vancouver, BC, Canada) for bibliographic identification at scale; the Unpaywall API (<https://api.unpaywall.org>, accessed on 25 June 2026; OurResearch) for open-access full-text retrieval; Google AI Studio (<https://aistudio.google.com>, accessed on 25 June 2026) for free-tier programmatic access to Gemma 3 27B and Gemini 2.5 Flash Lite during Studies A and B; OpenRouter (<https://openrouter.ai>, accessed on 25 June 2026) for free-tier access to NVIDIA Nemotron 3 Super 120B as Coder B in Study A; and Semantic Scholar (<https://api.semanticscholar.org>, accessed on 25 June 2026) for the citation-tracing step on the Joint Research Centre AI Watch literature. The authors thank the editor and the anonymous reviewers for their feedback on the manuscript.

Conflicts of Interest: Author Nusi Borovac was employed by the company Kenza. 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.

Abbreviations

The following abbreviations are used in this manuscript:

AGI	Artificial General Intelligence
AI	Artificial Intelligence
AI Act	EU Artificial Intelligence Act (Regulation (EU) 2024/1689)

CISR	Center for Information Systems Research (MIT)
DSR	Design Science Research
EIT	European Institute of Innovation and Technology
EU	European Union
FYP	Five-Year Plan
GenAI	Generative Artificial Intelligence
INSYTE	Intelligent Systems classification framework [28]
JRC	Joint Research Centre (European Commission)
KIC	Knowledge and Innovation Community (EIT)
L1–L6	Autonomy Levels (Assistive, Advisory, Supervisory, Delegated, Autonomous, Orchestrated)
MIT	Massachusetts Institute of Technology
NANDA	Networked Agents and Decentralized Architecture (MIT)
NIST	National Institute of Standards and Technology (United States)
NPC	National People’s Congress (China)
OSF	Open Science Framework
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ	Research Question
SEAI	Strategic Enterprise Artificial Intelligence [38]
SLR	Systematic Literature Review
SME	Small and Medium-Sized Enterprise
T1–T6	AI Types (Decision, Predictive, Generative, Conversational, Visual, Physical AI)

References

1. National People’s Congress of the People’s Republic of China. Outline of the 15th Five-Year Plan for National Economic and Social Development of the People’s Republic of China (2026–2030). Adopted by the Fourth Session of the Fourteenth National People’s Congress on 12 March 2026. On the “AI+” Action Plan Referenced Therein, See Also: State Council of the People’s Republic of China, Guideline on Deeply Implementing the “AI Plus” Initiative. 26 August 2025. Available online: https://english.www.gov.cn/policies/latestreleases/202508/27/content_WS68ae7976c6d0868f4e8f51a0.html (accessed on 25 June 2026).
2. European Parliament; Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Official Journal of the European Union, OJ L, 2024/1689, 12.7.2024, 2024. Available online: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (accessed on 25 June 2026).
3. European Commission. EU Launches InvestAI Initiative to Mobilise €200 Billion of Investment in Artificial Intelligence. Press Release IP/25/467, 2025. Available online: https://ec.europa.eu/commission/presscorner/detail/en/ip_25_467 (accessed on 25 June 2026).
4. European Commission. The AI Continent Action Plan. COM(2025) 165, 2025. Available online: <https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan> (accessed on 25 June 2026).
5. European Commission. Apply AI Strategy. COM(2025) 723 Final, 2025. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52025DC0723> (accessed on 25 June 2026).
6. Présidence de la République française. Clôture de la Première Journée du Sommet Pour l’Action sur l’IA. Sommet Pour l’Action sur l’intelligence artificielle, Grand Palais, Paris, 2025. Available online: <https://www.elysee.fr/emmanuel-macron/2025/02/10/cloture-de-la-premiere-journee-du-sommet-pour-laction-sur-lia> (accessed on 25 June 2026).
7. Challapally, A.; Pease, C.; Raskar, R.; Chari, P. *The GenAI Divide: State of AI in Business 2025*; Technical Report; MIT Project NANDA: Cambridge, MA, USA, 2025.
8. Gartner. Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. Press Release, 2025. Available online: <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027> (accessed on 25 June 2026).
9. Singla, A.; Sukharevsky, A.; Hall, B.; Yee, L.; Chui, M.; Balakrishnan, T. *The State of AI in 2025: Agents, Innovation, and Transformation*; Technical report; McKinsey & Company: New York, NY, USA, 2025.
10. Weill, P.; Woerner, S.L.; Sebastian, I.M. *Building Enterprise AI Maturity*; MIT CISR Research Briefing Vol. XXIV, No. 12; MIT Center for Information Systems Research: Cambridge, MA, USA, 2024.
11. Sadiq, R.B.; Safie, N.; Abd Rahman, A.H.; Goudarzi, S. Artificial Intelligence Maturity Model: A Systematic Literature Review. *PeerJ Comput. Sci.* **2021**, *7*, e661. [CrossRef] [PubMed]

12. Enholm, I.M.; Papagiannidis, E.; Mikalef, P.; Krogstie, J. Artificial Intelligence and Business Value: A Literature Review. *Inf. Syst. Front.* **2022**, *24*, 1709–1734. [\[CrossRef\]](#)
13. Mikalef, P.; Gupta, M. Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance. *Inf. Manag.* **2021**, *58*, 103434. [\[CrossRef\]](#)
14. Noy, S.; Zhang, W. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science* **2023**, *381*, 187–192. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Brynjolfsson, E.; Li, D.; Raymond, L.R. Generative AI at Work. *Q. J. Econ.* **2025**, *140*, 889–942. [\[CrossRef\]](#)
16. Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. GPTs Are GPTs: Labor Market Impact Potential of LLMs. *Science* **2024**, *384*, 1306–1308. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Collins, C.; Dennehy, D.; Conboy, K.; Mikalef, P. Artificial Intelligence in Information Systems Research: A Systematic Literature Review and Research Agenda. *Int. J. Inf. Manag.* **2021**, *60*, 102383. [\[CrossRef\]](#)
18. Newman, J. *A Taxonomy of Trustworthiness for Artificial Intelligence: Connecting Properties of Trustworthiness with Risk Management and the AI Lifecycle*; Cltc white paper series; UC Berkeley Center for Long-Term Cybersecurity: Berkeley, CA, USA, 2023.
19. Abercrombie, G.; Benbouzid, D.; Giudici, P.; Golpayegani, D.; Hernandez, J.; Noro, P.; Pandit, H.; Paraschou, E.; Pownall, C.; Prajapati, J.; et al. A Collaborative, Human-Centred Taxonomy of AI, Algorithmic, and Automation Harms. *arXiv* **2024**, arXiv:2407.01294.
20. Bagehorn, F.; Brimijoin, K.; Daly, E.M.; He, J.; Hind, M.; Garcés-Erice, L.; Giblin, C.; Giurgiu, I.; Martino, J.; Nair, R.; et al. AI Risk Atlas: Taxonomy and Tooling for Navigating AI Risks and Resources. *arXiv* **2025**, arXiv:2503.05780.
21. Mäntymäki, M.; Minkkinen, M.; Birkstedt, T.; Viljanen, M. Defining Organizational AI Governance. *AI Ethics* **2022**, *2*, 603–609. [\[CrossRef\]](#)
22. Birkstedt, T.; Minkkinen, M.; Tandon, A.; Mäntymäki, M. AI Governance: Themes, Knowledge Gaps and Future Agendas. *Internet Res.* **2023**, *33*, 133–167. [\[CrossRef\]](#)
23. Sheridan, T.B.; Verplank, W.L. *Human and Computer Control of Undersea Teleoperators*; Technical Report; Massachusetts Institute of Technology, Man–Machine Systems Laboratory: Cambridge, MA, USA, 1978.
24. Parasuraman, R.; Sheridan, T.B.; Wickens, C.D. A Model for Types and Levels of Human Interaction with Automation. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2000**, *30*, 286–297. [\[CrossRef\]](#)
25. Dellermann, D.; Ebel, P.; Söllner, M.; Leimeister, J.M. Hybrid Intelligence. *Bus. Inf. Syst. Eng.* **2019**, *61*, 637–643. [\[CrossRef\]](#)
26. Morris, M.R.; Sohl-Dickstein, J.; Fiedel, N.; Warkentin, T.; Dafoe, A.; Faust, A.; Farabet, C.; Legg, S. Position: Levels of AGI for Operationalizing Progress on the Path to AGI. In Proceedings of the Forty-First International Conference on Machine Learning (ICML 2024), Vienna, Austria, 21–27 July 2024; Volume 235.
27. Feng, K.J.K.; McDonald, D.W.; Zhang, A.X. Levels of Autonomy for AI Agents. *arXiv* **2025**, arXiv:2506.12469.
28. Porter, Z.; Calinescu, R.; Lim, E.; Hodge, V.; Ryan, P.; Burton, S.; Habli, I.; Lawton, T.; McDermid, J.; Molloy, J.; et al. INSYTE: A Classification Framework for Traditional to Agentic AI Systems. *ACM Trans. Auton. Adapt. Syst.* **2025**, *20*, 15. [\[CrossRef\]](#)
29. SAE Standard J3016_202104; Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International: Warrendale, PA, USA, 2021. [\[CrossRef\]](#)
30. EIT Climate-KIC. *Creation of a Taxonomy for the European AI Ecosystem*; Cross-Kic Activity “Innovation Impact Artificial Intelligence”; European Institute of Innovation and Technology: Budapest, Hungary, 2020.
31. Samoli, S.; López Cobo, M.; Delipetrev, B.; Martínez-Plumed, F.; Gómez, E.; De Prato, G. *AI Watch. Defining Artificial Intelligence 2.0: Towards an Operational Definition and Taxonomy for the AI Landscape*; Jrc Technical Report eur 30873 en; Publications Office of the European Union: Luxembourg, 2021. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Theofanos, M.; Choong, Y.Y.; Jensen, T. *AI Use Taxonomy: A Human-Centered Approach*; Nist ai 200-1; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2024. [\[CrossRef\]](#)
33. Davenport, T.H.; Ronanki, R. Artificial Intelligence for the Real World. *Harv. Bus. Rev.* **2018**, *96*, 108–116.
34. Davenport, T.H. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*; MIT Press: Cambridge, MA, USA, 2018.
35. Brynjolfsson, E.; Mitchell, T. What Can Machine Learning Do? Workforce Implications. *Science* **2017**, *358*, 1530–1534. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Brynjolfsson, E.; Mitchell, T.; Rock, D. What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Pap. Proc.* **2018**, *108*, 43–47. [\[CrossRef\]](#)
37. Herrmann, H. The Arcanum of Artificial Intelligence in Enterprise Applications: Toward a Unified Framework. *J. Eng. Technol. Manag.* **2022**, *66*, 101716. [\[CrossRef\]](#)
38. Bashir, J. Strategic Enterprise Artificial Intelligence (The Conceptual Hierarchical Framework). *Int. J. Bus. Manag. Stud.* **2024**, *5*, 131–138. [\[CrossRef\]](#)
39. Stein, H. Toward a Taxonomy of Generative AI Use Cases in Business Contexts: Integrating Complexity, Risk, and Strategy. *J. Comput. Sci.* **2026**, *96*, 102826. [\[CrossRef\]](#)

40. Sapkota, R.; Roumeliotis, K.I.; Karkee, M. AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. *Inf. Fusion* **2026**, *126*, 103599. [\[CrossRef\]](#)
41. Panigrahy, S. Multi-Agentic AI Systems: A Comprehensive Framework for Enterprise Digital Transformation. *J. Comput. Sci. Technol. Stud.* **2025**, *7*, 86–96. [\[CrossRef\]](#)
42. Arunkumar, V.; Gangadharan, G.R.; Buyya, R. Agentic Artificial Intelligence (AI): Architectures, Taxonomies, and Evaluation of Large Language Model Agents. *arXiv* **2026**, arXiv:2601.12560.
43. Fornasiero, R.; Kiebler, L.; Falsafi, M.; Sardesai, S. Proposing a Maturity Model for Assessing Artificial Intelligence and Big Data in the Process Industry. *Int. J. Prod. Res.* **2025**, *63*, 1235–1255. [\[CrossRef\]](#)
44. Iansiti, M.; Lakhani, K.R. *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*; Harvard Business Review Press: Boston, MA, USA, 2020.
45. Acemoglu, D.; Restrepo, P. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *J. Econ. Perspect.* **2019**, *33*, 3–30. [\[CrossRef\]](#)
46. ISO/IEC 42001:2023; Information Technology—Artificial Intelligence—Management System. International Organization for Standardization: Geneva, Switzerland, 2023.
47. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*; Nist ai 100-1; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023. [\[CrossRef\]](#)
48. Boston Consulting Group. *The AI Maturity Matrix: Which Economies Are Ready for AI?* Technical Report; BCG X/Boston Consulting Group: Boston, MA, USA, 2024.
49. Boston Consulting Group. *From Potential to Profit: Closing the AI Impact Gap*; Technical Report; Boston Consulting Group: Boston, MA, USA, 2025.
50. Deloitte. *The State of Generative AI in the Enterprise: Now Decides Next*; Technical Report; Deloitte AI Institute: New York, NY, USA, 2024; Quarterly Survey Series; Fourth-Wave Report 2024.
51. Accenture. *Technology Vision 2024: Human by Design—How AI Unleashes the Next Level of Human Potential*; Technical Report; Accenture Research: Dublin, Ireland, 2024.
52. Kanbach, D.K.; Heiduk, L.; Blüher, G.; Schreiter, M.; Lahmann, A. The GenAI Is Out of the Bottle: Generative Artificial Intelligence from a Business Model Innovation Perspective. *Rev. Manag. Sci.* **2024**, *18*, 1189–1220. [\[CrossRef\]](#)
53. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Shea, B.J.; Reeves, B.C.; Wells, G.; Thuku, M.; Hamel, C.; Moran, J.; Moher, D.; Tugwell, P.; Welch, V.; Kristjansson, E.; et al. AMSTAR 2: A Critical Appraisal Tool for Systematic Reviews That Include Randomised or Non-Randomised Studies of Healthcare Interventions, or Both. *BMJ* **2017**, *358*, j4008. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Whiting, P.; Savović, J.; Higgins, J.P.T.; Caldwell, D.M.; Reeves, B.C.; Shea, B.; Davies, P.; Kleijnen, J.; Churchill, R. ROBIS: A New Tool to Assess Risk of Bias in Systematic Reviews Was Developed. *J. Clin. Epidemiol.* **2016**, *69*, 225–234. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Haddaway, N.R.; Grainger, M.J.; Gray, C.T. Citationchaser: A Tool for Transparent and Efficient Forward and Backward Citation Chasing in Systematic Searching. *Res. Synth. Methods* **2022**, *13*, 533–545. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *MIS Q.* **2004**, *28*, 75–105. [\[CrossRef\]](#)
58. 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. [\[CrossRef\]](#)
59. Nickerson, R.C.; Varshney, U.; Muntermann, J. A Method for Taxonomy Development and Its Application in Information Systems. *Eur. J. Inf. Syst.* **2013**, *22*, 336–359. [\[CrossRef\]](#)
60. Drljevic, N.; Arias Aranda, D.; Stantchev, V. An Integrated Adoption Model to Manage Blockchain-Driven Business Innovation in a Sustainable Way. *Sustainability* **2022**, *14*, 2873. [\[CrossRef\]](#)
61. Eurostat. *Use of Artificial Intelligence in Enterprises*; Eurostat Statistics Explained, 2024 Survey Results; Eurostat: Luxembourg, 2025.
62. Arroyabe, M.F.; Arranz, C.F.A.; Fernandez De Arroyabe, I.; Fernandez de Arroyabe, J.C. Analyzing AI Adoption in European SMEs: A Study of Digital Capabilities, Innovation, and External Environment. *Technol. Soc.* **2024**, *79*, 102733. [\[CrossRef\]](#)

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.