

Review

Digital Twins and Artificial Intelligence for HAZOP Enhancement in Process Safety: A Critical Literature Review

Feras Alrowaie 

Chemical Engineering Department, Jubail Industrial College, Royal Commission for Jubail & Yanbu,
Jubail Industrial City 35718, Saudi Arabia; rowaie_fa@rcjy.edu.sa

Abstract

Conventional hazard and operability (HAZOP) studies remain central to process safety management, but their periodic, document-centered implementation and dependence on expert judgment limit their ability to track dynamic operational risk across the full plant lifecycle. This critical literature review, drawing on 113 sources, examines how digital twin (DT) and artificial intelligence (AI) technologies may augment HAZOP practice, advancing an augmentation principle: DT–AI should strengthen expert-led hazard analysis as a decision-support layer, not replace human judgment. The review maps conventional HAZOP limitations to four complementary technology pathways: AI-assisted knowledge capture and deviation reasoning; DT-based hazard monitoring and early warning; hybrid physics–data modeling for predictive safety; and explainable AI for operator trust and regulatory acceptance. Four representative industrial case studies illustrate current practice, and a structured barrier analysis identifies why many pilots stall before reaching sustained deployment. These pathways offer strong potential to improve HAZOP coverage, traceability, operational relevance, and lifecycle learning, but a study-level evidence assessment finds that no reviewed study demonstrates long-term, multi-site validation of an integrated DT–AI–HAZOP system; most implementations remain at conceptual, prototype, or limited pilot levels. Key failure modes, including model drift, sensor faults, large language model hallucination, and automation complacency, are examined alongside a thirteen-item implementation risk register spanning technical, organizational, economic, security, and scalability concerns. The review proposes a phased industrial implementation roadmap, a lifecycle-oriented safety-management framework, and a three-horizon research agenda for advancing toward reliable, validated HAZOP-informed digital twin systems.

Keywords: digital twins; artificial intelligence; HAZOP; process safety; hazard monitoring; ontology; knowledge graph; explainable AI; hybrid modeling; lifecycle safety management



Academic Editors: Sheng Du, Li Jin,
Pan Yu and Hao Liu

Received: 9 June 2026

Revised: 12 July 2026

Accepted: 15 July 2026

Published: 17 July 2026

Copyright: © 2026 by the author.

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.

1. Introduction

1.1. Motivation: HAZOP in a Dynamic Operating Environment

On 23 March 2005, an explosion at the BP Texas City refinery killed 15 workers and injured more than 180 others, making it one of the deadliest industrial accidents in recent United States history. The investigation by the U.S. Chemical Safety and Hazard Investigation Board (CSB) concluded that the accident involved the overfilling of a raffinate splitter tower during start-up. More importantly, the event reflected a deeper pattern of operational drift, weak learning from previous incidents, and insufficient reassessment of hazards as plant conditions changed over time [1]. Only a few months later, the Buncefield fuel storage terminal explosion in the United Kingdom was triggered by a tank-overfilling

event that was not effectively detected or stopped by existing protection layers. The explosion caused widespread damage across the surrounding industrial area and led to a major reassessment of fuel-storage safety, alarm management, and overfill-prevention practices [2,3]. Although the two incidents occurred in different sectors and operating contexts, they point to the same underlying concern: major hazards may evolve gradually during operation, while formal hazard reviews are often conducted only periodically.

This concern is not limited to BP Texas City or Buncefield. Investigations of major process safety events frequently show that serious accidents are rarely caused by a single technical failure. They often emerge from a combination of degraded equipment, weak barrier management, organizational decisions, and deviations from the original design intent. The Macondo blowout, the Chevron Richmond refinery fire, and the IOC Jaipur terminal fire all illustrate how abnormal conditions and safety-critical weaknesses can accumulate over time before becoming visible as major accidents [4–7]. These examples do not suggest that HAZOP is ineffective. Rather, they highlight a limitation in how HAZOP is commonly applied. In many facilities, HAZOP remains an episodic, design-stage or revalidation activity. This makes it difficult for the method to continuously reflect operational drift, equipment aging, process modifications, changing feedstocks, or emerging abnormal situations between formal review cycles [8–10]. Critiques of HAZOP limitations [11] and guidance on revalidation [12] reinforce this concern.

Recent advances in industrial sensing, data infrastructure, computational modeling, and AI-based analytics now make it more realistic to connect process safety knowledge with live operational data. At the same time, many process industries are becoming more instrumented, data-rich, and connected. Real-time sensor networks, industrial artificial intelligence, edge and cloud computing, and digital twin platforms are creating new possibilities for monitoring plant behavior continuously rather than relying only on periodic review and manual interpretation [13,14]. Digital twin platforms in particular offer structured representations of physical assets that can be continuously updated with operational data [15,16]. In parallel, recent work on ontology-based HAZOP automation, process safety knowledge graphs, and natural language processing shows that parts of hazard identification and safety knowledge management can be supported by computational tools [17–19]. These developments create an important opportunity: not to replace HAZOP, whose structured and expert-driven reasoning remains central to process safety, but to extend it beyond the design and revalidation stage. A digitally enhanced HAZOP framework could help compare current plant behavior with design intent, detect emerging deviations, update safety knowledge, and support more timely decision making during operation.

This review therefore examines how digital twin and artificial intelligence technologies can support the evolution of HAZOP from a mainly periodic hazard-review method toward a more dynamic, lifecycle-oriented safety management approach. The aim is not to present digitalization as a simple solution to process safety problems, but to critically evaluate where these technologies can add value, where their limitations remain, and what technical and organizational barriers must be addressed before they can be adopted reliably in high-hazard process industries. This conceptual transition is summarized in Figure 1. Conventional HAZOP remains the foundation of expert-led process hazard analysis, but its periodic and document-centered implementation creates limitations that can be addressed through four complementary DT–AI pathways: AI-assisted HAZOP, digital twin-based monitoring, hybrid physics–data models, and explainable AI.

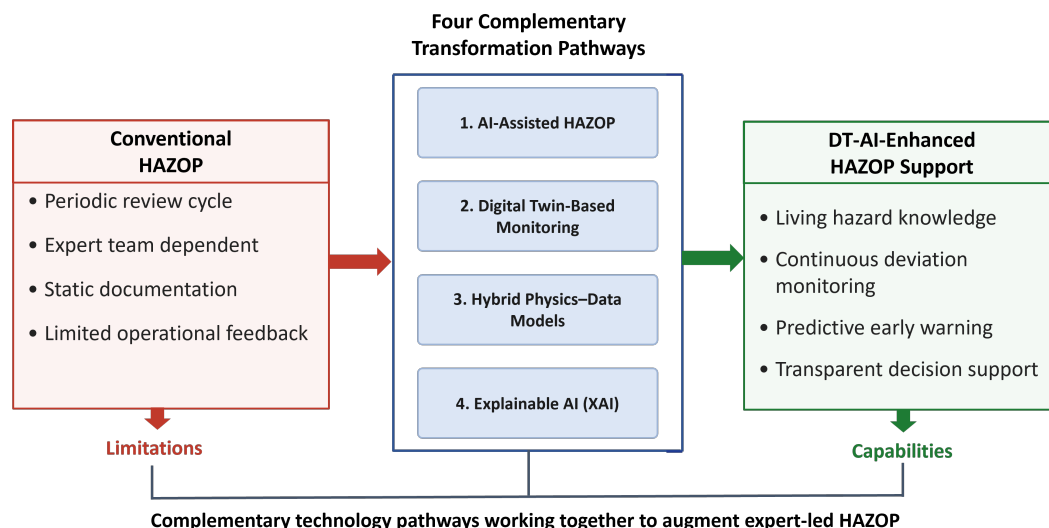


Figure 1. Evolution from conventional HAZOP to DT-AI-enhanced HAZOP support through four complementary transformation pathways.

1.2. HAZOP as a Foundation of Process Safety

Hazard and operability (HAZOP) studies have been the central method for systematic hazard identification in the chemical and process industries since their development within Imperial Chemical Industries in the 1960s [10,20]. A multidisciplinary team divides the process into nodes and applies guide words (“NO”, “MORE”, “LESS”, “REVERSE”, “OTHER THAN”) to process parameters, identifying credible deviations, causes, consequences, safeguards, and recommendations [8–10]. The method’s strength lies in the balance it strikes between structured, systematic questioning and expert judgment: the guide-word approach reduces the chance that important deviations are missed, while the study quality still depends heavily on team experience, documentation accuracy, and participants’ ability to recognize how abnormal conditions develop in practice [8,10,11].

HAZOP is also deeply embedded in process safety regulation: it underpins process hazard analysis requirements under the U.S. OSHA Process Safety Management standard [21], the EU Seveso III Directive [22], and the UK COMAH Regulations 2015 [23], and is codified in the CCPS *Guidelines for Hazard Evaluation Procedures* and IEC 61882:2016 [8,9]. This regulatory and standards foundation has made HAZOP a common language for hazard identification across companies, regulators, and engineering teams; readers seeking a comprehensive treatment of the method itself should consult IEC 61882 and the CCPS guidelines directly.

Despite this strong foundation, HAZOP is not a complete solution by itself. Literature reviews and critiques note recurring limitations related to team dependence, time requirements, documentation quality, and the difficulty of capturing dynamic operational changes within a periodic review process [10–12], limitations that are especially consequential in modern plants, where operating conditions, control systems, maintenance status, feedstocks, and equipment health can change continuously between formal review cycles. These considerations motivate the research gap discussed in the next section: how HAZOP can be digitally supported while preserving its structured, expert-led, and auditable character.

1.3. Research Gap and Novel Contribution

1.3.1. The Research Gap

A growing body of research has examined how artificial intelligence and digital twin technologies can support different aspects of process safety. Some studies have focused on ontology-based and computer-aided HAZOP automation [17,18], while others have explored knowledge graph construction for process safety knowledge integration [19], digital twin models for process-plant risk monitoring [16], safety assurance for machine learning systems [24], and interpretable machine learning for safety-related decision support [25,26]. These contributions are important because they show that digital technologies can support specific parts of hazard identification, safety monitoring, knowledge management, and decision making.

However, the literature remains fragmented. Existing studies tend to address individual technologies or isolated use cases, rather than presenting a unified view of how digital twins, AI-assisted HAZOP, hybrid physics–data models, and explainable AI could work together to strengthen HAZOP practice. In particular, there is still a need for a critical synthesis that organizes these developments into technology pathways aligned with the practical limitations of conventional HAZOP, compares their evidence maturity and implementation requirements, and examines the gap between laboratory demonstrations and industrial deployment. There is also limited discussion of the regulations, human factors, and organizational issues that determine whether these tools can be adopted safely in real process facilities.

Table 1 summarizes representative prior studies that have informed this field and highlights the specific gaps that motivate the present review.

Table 1. Representative prior studies informing DT–AI–HAZOP integration and the research gaps addressed in this review.

Representative Prior Study	Primary Focus	Key Research Gap
Schwalbe and Schels [24]	Lifecycle- and certification-oriented safety assurance methods for ML-based systems.	Does not address HAZOP-specific integration or DT–ML coupling for process safety.
Bevilacqua et al. [16]	Digital twin reference model for operator-risk prediction and prevention in process plants.	Does not specifically develop AI-assisted HAZOP or XAI requirements for HAZOP decision support.
Barredo Arrieta et al. [25]	General XAI concepts, taxonomies, opportunities, and challenges.	Provides a broad XAI foundation but does not address HAZOP/PHA workflows, digital twin integration, or HAZOP-specific interpretability requirements.
Single et al. [17,18]	Ontology-based, computer-aided generation of HAZOP worksheets and hazard scenarios.	Does not address operational DT integration, hybrid physics–data modeling, or XAI for HAZOP decision support.
Mao et al. [19]	Process safety knowledge graph and domain-ontology development using sources including HAZOP reports.	Does not address digital twin architectures, hybrid physics–data models, or XAI requirements for regulatory acceptance.
Mohseni et al. [26]	Multidisciplinary framework for the design and evaluation of XAI systems.	Does not specifically address HAZOP automation, DT–HAZOP integration, or HAZOP-domain knowledge representation.

Table 2 compares this review with selected foundational and related studies across seven dimensions: AI-assisted HAZOP, digital twin coverage, hybrid physics–data modeling, explainable AI, human factors, standards and regulatory considerations, and technology-maturity assessment. In the AI-assisted HAZOP column, partial coverage denotes computational formalization of HAZOP-relevant knowledge without a complete AI-assisted workflow. Throughout this review, Δ denotes partial or intermediate coverage and is distinct from the tilde ($\sim$), which is used elsewhere to mean “approximately”.

Table 2. Scope-based comparison of this review with selected foundational and related works. Symbols indicate substantive topic coverage rather than methodological quality, evidentiary strength, or superiority: $\checkmark$ = central coverage; Δ = partial or indirect coverage; $\times$ = not substantively addressed.

Review	AI-Assisted HAZOP	Digital Twin	Physics–Data Models	XAI	Human Factors	Standards/Regulation	Technology Maturity
Single et al. [17,18]	$\checkmark$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Bevilacqua et al. [16]	$\times$	$\checkmark$	$\times$	$\times$	Δ	$\times$	$\times$
Barredo Arrieta et al. [25]	$\times$	$\times$	$\times$	$\checkmark$	Δ	Δ	$\times$
Mao et al. [19]	Δ	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Schwalbe & Schels [24]	$\times$	$\times$	$\times$	Δ	$\times$	$\checkmark$	Δ
Mohseni et al. [26]	$\times$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\times$
This review	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

1.3.2. Novel Contributions of This Review

This review contributes to the field in five main ways. For each, the provenance of the contribution is stated explicitly: what is drawn from prior literature, what is the author’s own synthesis or interpretation of previously separate evidence, and what is newly proposed.

1. **Integrated critical synthesis (newly proposed organizing framework).** It provides an integrated critical synthesis of four technology pathways: AI-assisted HAZOP, digital twin-based monitoring, hybrid physics–data modeling, and explainable AI. This four-pathway organization around HAZOP’s specific limitations is not present as such in any single prior review (Table 1 documents the specific gap each existing review leaves); the individual pathways themselves are drawn from the established literature, but their integration around a common problem framing is new to this review.
2. **HAZOP-limitation-to-capability mapping (newly proposed structuring device).** It links key limitations of conventional HAZOP, including episodic review, dependence on expert judgment, static risk representation, limited use of operational data, and knowledge loss between studies, to specific digital twin and AI capabilities. This mapping is the author’s own structuring of previously disconnected literature and is intended to provide a practical basis for technology selection, implementation planning, and future research priorities.
3. **The augmentation principle (adopted from prior literature, newly applied as a consistent cross-pathway framework).** The principle that AI should act as decision support rather than a HAZOP replacement is present in the prior literature on safety-critical AI; what is new here is its consistent application as the organizing lens across all four technology pathways, rather than being confined to a single pathway or use case.

4. **Cross-pathway evidence-quality assessment (author’s own synthesis).** It evaluates evidence quality and technology maturity across the four pathways by distinguishing between conceptual proposals, simulation-based demonstrations, prototype implementations, and validated industrial applications (Table 3, Supplementary Table S1). This synthesis draws on evidence reported in the individual primary studies, but the cross-pathway comparison and the study-level evidence maturity assessment are the author’s own analysis, and not present in any cited source.
5. **Integrated adoption-readiness treatment (newly proposed synthesis).** It brings together technical, organizational, and human-centered adoption issues in one critical discussion, including explainability, regulatory acceptance, failure modes (Section 8), implementation barriers, a structured risk register, a phased implementation roadmap, and ethical considerations. Individual elements are informed by the cited literature, but their integration into a single, HAZOP-specific adoption framework is newly proposed by this review.

Table 3. Qualitative maturity comparison of the four DT-AI pathways for HAZOP enhancement.

Pathway	Current Maturity and Evidence Base	Main Gap Before Wider Industrial Adoption
AI-assisted HAZOP	Moderate research maturity. Evidence includes rule-based systems, ontology-based HAZOP, knowledge graphs, NLP tools, and LLM-assisted drafting. Most applications remain expert-support or prototype tools.	Industrial workflow evidence is needed, including expert-reviewed outputs, traceable AI-generated suggestions, version control, and governance procedures for formal HAZOP use.
Digital twin-based monitoring	Moderate and growing maturity. Evidence includes process monitoring, dynamic simulation, operator-support prototypes, and equipment-level safety applications. Many systems are not yet explicitly HAZOP-informed.	Closer linkage is needed between live plant data, structured HAZOP knowledge, safeguard status, model outputs, and operator decision support.
Hybrid physics–data models	Well-developed methodological foundation, but variable safety-deployment maturity. Evidence includes process fault diagnosis, physics-informed learning, predictive monitoring, and model-based state estimation.	Validation is needed under abnormal, transient, degraded, and rare-event conditions, together with uncertainty communication, model-drift monitoring, and model-governance procedures.
Explainable AI	Conceptually mature, but process safety-specific application remains developing. Evidence includes general XAI methods, AI safety-assurance frameworks, feature-attribution methods, and explanation tools.	Explanations should be expressed in HAZOP language and integrated with operator trust, auditability, regulatory review, human-in-the-loop decision making, and lifecycle safety learning.

Note: The maturity descriptions are qualitative author assessments based on the type, scale, validation setting, and deployment evidence reported in the reviewed literature. They are pathway-level syntheses and should not be interpreted as formal TRL assignments for every method or implementation. Detailed component-level TRL estimates are provided in Supplementary Table S11.

1.3.3. Objectives

This review critically evaluates how digital twin and artificial intelligence technologies can support the enhancement of conventional HAZOP and contribute to more dynamic process safety management. The specific objectives are as follows:

1. To characterize the main technological approaches for digital twin and AI-enhanced HAZOP across four complementary pathways.
2. To assess the available validation evidence, industrial implementation status, and technology readiness level of these approaches.

3. To compare the four pathways as complementary components of a lifecycle-oriented safety-management framework.
4. To identify potential failure modes, implementation barriers, and risk-control requirements associated with digital twin and AI adoption in process safety.
5. To provide practical guidance for phased industrial adoption, sector-specific considerations, and future research directions.

1.3.4. Scope and Perspective

This review is a thematic critical literature review rather than a PRISMA-style systematic review. The detailed search strategy, inclusion and exclusion criteria, selection process, and synthesis approach are described in Section 1.4. The review focuses specifically on digital twin and AI technologies in the context of HAZOP enhancement and process safety management for continuous and batch chemical processes. It does not attempt to survey the broader fields of AI, digital twins, or machine learning in full. The discussion gives attention not only to technical capability, but also to implementation realism, human oversight, regulatory acceptance, explainability, and organizational readiness.

1.3.5. Manuscript Organization

The remainder of this review is organized as follows. Section 1.4 describes the review methodology, including the search strategy, inclusion and exclusion criteria, selection process, and synthesis approach. Section 2 establishes the conceptual foundations required to understand digital twin and AI-enhanced HAZOP, including HAZOP knowledge structure, digital twin architecture for process safety, AI methods relevant to HAZOP enhancement, and the integration logic that connects static hazard review to lifecycle hazard intelligence. Sections 3–6 examine the four major technology pathways: AI-assisted HAZOP, digital twin architectures for dynamic hazard monitoring, hybrid physics–data models for predictive safety analytics, and explainable AI for operator trust and regulatory acceptance. Section 7 reviews industrial applications and technology maturity. Section 8 presents the critical discussion, comparative analysis, research gaps, and future research agenda. Section 8.13 discusses ethical and societal considerations. Section 8.14 covers verification, validation, and functional safety standards. Finally, Section 9 presents the conclusions and recommendations.

1.4. Review Methodology

This review follows a thematic critical synthesis methodology rather than a formal PRISMA-style systematic review, but the literature identification and selection process was documented transparently to support traceability and partial repeatability of the search.

1.4.1. Search Strategy and Selection Criteria

The literature used was identified through searches of Scopus and Google Scholar, supplemented by Web of Science, SciSpace, backward and forward citation tracing, and relevant process safety and process systems engineering conference literature. Scopus was queried using field-tagged TITLE-ABS-KEY() syntax to restrict retrieval to titles, abstracts, and author keywords; Google Scholar was queried using an equivalent plain-text Boolean string with a custom date filter, searching across full text and metadata. The search covered publications from 2000 to 2026, with primary emphasis on 2015 onward to reflect the period of substantive growth in digital twin and AI-assisted process safety research.

Search terms were organized into five Boolean query groups (A–E; see Supplementary Table S9) corresponding to the four technology pathways examined in this review: (A) an overall corpus anchor combining HAZOP and process hazard analysis terms with any digital twin, AI, or XAI term; (B) the AI-assisted HAZOP pathway, combining HAZOP and

process hazard analysis with ontologies, knowledge graphs, natural language processing, large language models, and machine learning; (C) the digital twin monitoring pathway, combining HAZOP and process safety with digital twin and cyber–physical system terms; (D) the hybrid physics–data modeling pathway, combining process safety, fault detection, and anomaly detection with physics-informed neural networks (PINNs) and hybrid model terms; and (E) the explainable AI and operator-trust pathway, combining process safety, HAZOP, and fault detection with XAI, SHAP, LIME, operator trust, and interpretability terms. Full query strings and indicative per-database hit counts from the revision-stage verification search are reported in Supplementary Table S9.

Sources were included when they addressed at least one of the four technology pathways as a primary or substantive secondary focus; provided sufficient technical detail to assess the method, its validation evidence, or its industrial applicability; and reported an original method, implementation, case study, or empirical evaluation (reviews were included only when they synthesized pathway-relevant evidence not otherwise captured). Sources were excluded when they addressed HAZOP, process safety, or AI only tangentially without a methodological or applied contribution to a pathway; lacked sufficient documentation to assess validity or scope (e.g., extended abstracts, posters, or papers without a methods or results section); or were purely theoretical, with no connection to process safety application or validation.

Candidate records identified through these searches were screened against the criteria given above by a single reviewer, first by title and abstract and then by full text; because screening was conducted by a single reviewer, no inter-rater reliability measure was computed. For the included sources, extracted information covered model or method type, application domain and scale, validation approach, reported performance metrics, TRL indicators where inferable, and stated limitations. No formal quality-scoring instrument was applied; evidence strength was instead assessed qualitatively, giving greater weight to studies with transparent validation procedures, industrial- or pilot-scale evidence, clear reporting of application scope, and findings independently replicated or corroborated across sources [27–29]. The synthesis was conducted thematically, structured around the four technology pathways; Section 1.4.2 reports the exact final corpus size and the availability of retained selection information.

1.4.2. Literature Corpus Summary

To make the reviewed literature transparent, the final corpus comprises 113 sources cited in this review, based on an exact count of the reference list. Table 4 provides an author-classified distribution across four primary thematic categories and a cross-cutting category covering HAZOP-foundational material, general process safety background, human factors, standards, and governance. Each reference was assigned to one category according to its dominant contribution for mutually exclusive counting, while secondary relevance to other categories was acknowledged. These descriptive counts characterize the cited corpus and should not be interpreted as the complete size of the wider literature in each field. Because this is a thematic rather than PRISMA-style systematic review, a PRISMA flow diagram was not used. Instead, Supplementary Table S9 provides a revision-stage verification search log: the five Boolean queries were re-executed on Scopus and Google Scholar in July 2026, with per-query hit counts recorded. Web of Science and SciSpace were consulted during the broader literature-identification process, but equivalent numerical counts are not reported because query syntax and indexing coverage were not directly comparable. Raw database hit counts are therefore verification-search results, not counts of unique screened records.

The original review was developed iteratively and did not retain a deduplicated screening ledger. Consequently, exact historical counts for unique records screened at title/abstract level, full texts assessed, and records excluded cannot be reconstructed reliably. Table 5 distinguishes the retained information from quantities that cannot be reconstructed and separates the revision-stage verification searches from the historical screening process.

Although complete historical screening counts are unavailable, the final cited corpus contains 113 identifiable references and forms the basis for the thematic synthesis summarized in Table 4.

Table 4. Distribution of the 113 cited sources across author-assigned primary thematic categories using a single-category classification.

Primary Thematic Category	Cited Sources, <i>n</i>	Main Evidence Types	Publication Year Range
AI-assisted HAZOP and safety-knowledge formalization (expert systems, ontologies, knowledge graphs, NLP, ML, and LLMs)	26	Academic case studies, computational demonstrations, and review articles	1994–2026
Digital twin architectures for hazard monitoring	15	Simulations, academic prototypes, limited field pilots, and selected industrial cases	2016–2025
Hybrid physics–data predictive models (PINNs, physics-informed learning, hybrid soft sensors, and uncertainty-aware methods)	15	Simulations, laboratory studies, and limited pilot studies	2003–2025
Explainable AI and human-centered assurance (interpretability, operator trust, safety assurance, and governance)	21	Framework papers, empirical studies, standards, and governance literature	2016–2025
Cross-cutting foundations (HAZOP, incident evidence, standards, ethics, and general AI)	36	Standards, guidelines, incident reports, foundational papers, and general reviews	1992–2025
Total (exact)	113	Mixed	1992–2026

Note: Categories are mutually exclusive—one primary category per reference by dominant contribution—and sum to 113; some references also contribute secondarily to others. See Supplementary Table S9 for the verification search log.

Table 5. Available literature-selection records and final cited corpus.

Information Reported	Count	Interpretation
Revision-stage verification searches	See Supplementary Table S9	Post hoc Scopus and Google Scholar searches conducted in July 2026; per-query results overlap and were not deduplicated or used as historical screening counts.
Unique records screened by title and abstract	NR	The original iterative review did not retain a deduplicated screening ledger.
Full-text reports assessed	NR	The historical number of full-text reports assessed cannot be reconstructed reliably.
Full-text reports excluded	NR	Historical exclusion counts and source-level exclusion reasons were not retained.
References in the final cited corpus	113	Exact number of cited references appearing in the compiled bibliography and used in the revised manuscript.

Note: NR indicates information that cannot be retrospectively reconstructed from the original thematic-review records.

2. Conceptual Foundations for Digital Twin- and AI-Enhanced HAZOP

Before examining specific technology pathways, this section clarifies the conceptual foundations that allow HAZOP, digital twins, and artificial intelligence to be discussed as connected parts of one safety-support framework rather than as separate tools. HAZOP captures expert knowledge about deviations, causes, consequences, safeguards, and required actions; digital twins provide a continuously updated representation of the operating process by linking process models with live plant data; AI provides computational support for knowledge extraction, anomaly recognition, prediction, diagnosis, and explanation. The central question is not simply how to add AI or digital twin tools to HAZOP, but how static hazard knowledge can be transformed into dynamic, explainable, lifecycle-oriented safety intelligence.

2.1. HAZOP as Structured Safety Knowledge

A conventional HAZOP worksheet organizes process safety reasoning through a structured sequence of nodes, parameters, guide words, deviations, causes, consequences, safeguards, and recommendations, as codified in the CCPS and IEC HAZOP guidance [8,9], synthesized in the Dunj  et al. literature review [10], and documented in standard process safety references [30,31]. This sequence allows HAZOP knowledge to be traced, reviewed, critically checked, updated, and encoded in computational systems. Guide words (“NO”, “MORE”, “LESS”, “REVERSE”, “OTHER THAN”) combined with process parameters (flow, pressure, temperature, level, composition) generate deviations; for each meaningful deviation, the expert team identifies credible causes, evaluates consequences, reviews existing safeguards, and proposes recommendations where protection is inadequate [8,10,30]. This structure is one of the main reasons HAZOP remains relevant in a digital environment: unlike a generic monitoring system that detects a statistically abnormal variable without explaining its safety relevance, a HAZOP-informed system connects abnormal behavior to a known deviation, credible cause, possible consequence, and expected safeguard, providing the semantic safety layer that a meaningful digital twin requires.

Recent work on computer-aided hazard analysis [32,33], ontology-based HAZOP [17,18], and process safety knowledge graphs [19] shows that HAZOP worksheets can be transformed from narrative documentation into structured knowledge models supporting retrieval, consistency checking, and reuse, not to remove expert judgment, but to make it more traceable and connectable to operational data. Interpreted in this way, a HAZOP worksheet becomes a structured safety knowledge base: the deviation column informs abnormal-condition monitoring, the cause column supports fault diagnosis, the consequence column supports risk prioritization [34], and the safeguard column supports protection-layer verification. Figure 2 illustrates this formalization into a machine-readable safety knowledge base connected to DT-AI functions for monitoring, analytics, decision support, and lifecycle learning.

2.2. From Static HAZOP Worksheets to Dynamic Lifecycle Safety Intelligence

Conventional HAZOP is organized around discrete study events (design, major modifications, periodic revalidation), which creates an auditable, multidisciplinary, regulation-compliant record but means the worksheet can drift from actual operating conditions over time as equipment fouls, catalysts age, sensors drift, feedstocks vary, and modifications accumulate [11,12,35–37]. Digital twin technology offers a way to narrow this gap. The foundational digital twin concept, a virtual representation kept continuously synchronized with a physical asset through live data [15], has since been adapted specifically for process safety applications: reference architectures for operator-risk prevention [16], safety analysis and emergency management [38], integrated plant monitoring [39], and process safety management in oil-and-gas and other process-industry operating units [40–44], with broader chemical-industry adoption surveyed in [45]. A system connected to HAZOP knowledge can evaluate whether current behavior is approaching a known deviation, e.g., relating deteriorating heat-transfer performance to a shrinking margin against a high-temperature deviation, or rising column differential pressure to flooding-related consequences.

This transforms HAZOP from a static worksheet into part of a living safety knowledge system: the original expert analysis still defines credible deviations, causes, consequences, and safeguards, but operational data can help evaluate whether these scenarios are becoming more likely and whether the study’s assumptions still hold. The digital twin does not replace HAZOP; it provides an operational layer keeping HAZOP knowledge connected to plant behavior. This extends beyond simple alarm generation: rather than a variable crossing a fixed limit, a HAZOP-informed digital twin can provide richer context (which

deviation is developing, which causes are plausible, which safeguards should be checked, how much time remains), supporting more explainable early warning. The extent to which this has been realized in practice remains uneven. Later sections assess where the evidence is from industry versus at the conceptual or prototype-level.

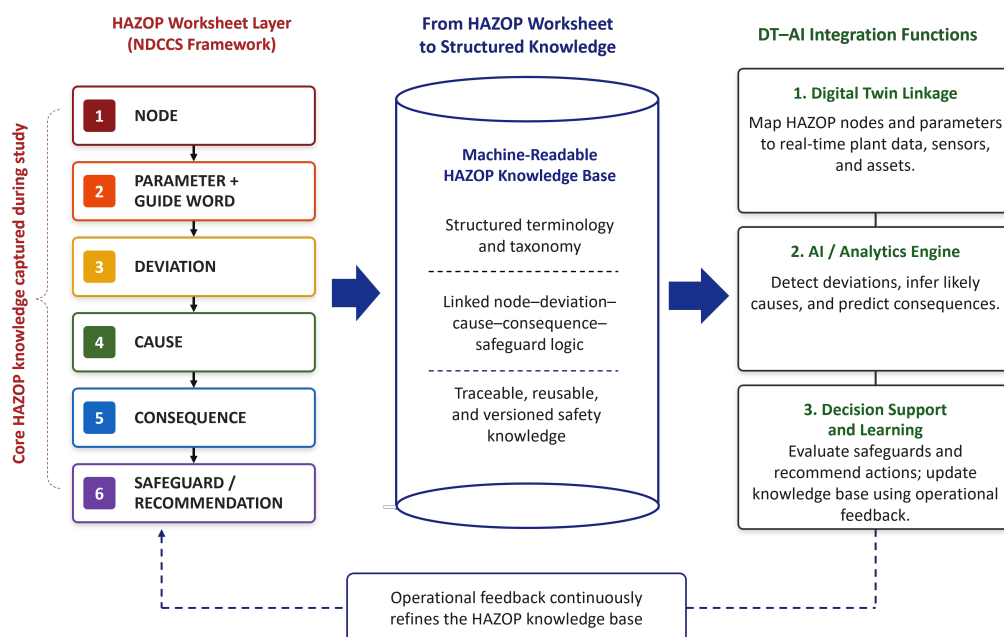


Figure 2. Structured HAZOP knowledge representation for DT-AI integration. Conventional HAZOP worksheet elements can be interpreted as a node–deviation–cause–consequence–safeguard framework that supports knowledge formalization, digital twin linkage, AI analytics, decision support, and continuous safety learning.

2.3. Digital Twin Architecture for Process Safety Applications

If HAZOP provides the structured safety knowledge, the digital twin provides the operational environment in which that knowledge can be compared with plant behavior [15,16,38]; its value in process safety depends on connecting live data, process models, equipment status, safety logic, and operator decision support into one coherent architecture, not merely visualization or simulation. The five-layer architecture developed in Section 4.2, comprising physical asset, data acquisition, model, analytics, and decision-support layers, clarifies which functional layers are required when HAZOP knowledge is to be connected to monitoring, prediction, diagnosis, and decision support. The key distinguishing feature from a generic digital twin is that HAZOP deviations define what abnormal situations should be monitored, HAZOP causes guide diagnostic feature selection, HAZOP consequences support risk prioritization, and HAZOP safeguards provide the basis for protection-layer verification. Digital twin maturity in process safety varies significantly. Applications span operator-risk prevention [16], dynamic safety analysis [38], hydrogen refueling station safety [46], and oil-and-gas and petrochemical process safety management [40–44], but most systems focus on visualization, anomaly detection, or equipment health rather than explicitly connecting to HAZOP’s deviation–cause–consequence–safeguard logic, motivating the HAZOP-informed architecture developed in this review.

2.4. Artificial Intelligence Methods Relevant to HAZOP Enhancement

AI is not a single method; it is a family of computational approaches that support different parts of the HAZOP lifecycle, with different maturity levels and safety risks [13,14,47]. Rule-based systems can encode explicit deviation–cause–consequence logic and remain auditable, but are limited by rule-base coverage and maintainability [32,33]. Ontologies

and knowledge graphs convert HAZOP information into reusable machine-readable relationships, supporting retrieval, consistency checking, and linkage to operational data, although they require substantial knowledge-engineering effort [17–19]. Natural language processing and LLM-based methods [27,48–50] can extract or draft structured HAZOP content from worksheets, incident reports, and procedures, but fluent text does not guarantee correct causal reasoning, adequate coverage, or physical consistency.

For operational digital twin use, machine learning models can identify multivariable abnormal patterns, Bayesian and dynamic Bayesian networks can update cause–consequence likelihoods as evidence changes [34,38], and hybrid physics–data models can reduce the extrapolation risk of purely data-driven models by embedding process constraints, an approach established by the foundational physics-informed neural network framework and subsequent reviews [51,52] and applied to fault detection in nuclear power valves [53], industrial wastewater treatment [54], and predictive process control [55]. In all cases, explainability is essential: outputs must show why a recommendation was made, which variables contributed, and whether the situation lies within the validated operating domain [24–26]. This requirement is meaningful only when expressed in HAZOP terms (deviation, cause, consequence, safeguard, action), not as a cosmetic layer.

AI is most credible when it augments expert-led analysis, knowledge management, monitoring, or diagnosis, and least credible when presented as a replacement for expert judgment: the augmentation principle requires that responsibility for safety-critical decisions remain with qualified human professionals and established safety management systems. Table 6 summarizes the complementary roles of HAZOP, digital twins, AI analytics, and explainability in this lifecycle safety-intelligence framework.

Table 6. Conceptual roles of HAZOP, digital twins, AI analytics, and explainability in a lifecycle safety-intelligence framework.

Element	Primary Contribution	Role in HAZOP Enhancement
HAZOP	Structured expert safety knowledge	Defines deviations, causes, consequences, safeguards, and recommendations.
Digital twin	Live operational and model-based process representation	Connects HAZOP knowledge to current plant behavior, equipment condition, and evolving safety margins.
AI analytics	Pattern recognition, prediction, and diagnosis	Supports knowledge extraction, abnormal-condition detection, cause ranking, predictive warning, and decision support.
Explainable AI	Transparent and auditable reasoning support	Helps operators and engineers understand why an alert or recommendation was generated and whether it is credible.

2.5. Integration Logic: From HAZOP Knowledge to Lifecycle Hazard Intelligence

HAZOP, digital twins, and AI each contribute a different capability: structured safety knowledge; operational context; and computational support for pattern recognition, prediction, diagnosis, and explanation, respectively. The central question is how to integrate them without weakening the expert judgment and accountability that make HAZOP valuable. This integration is best viewed as a progression from static hazard identification to lifecycle hazard intelligence, organized around four functions [28]: *identify* (conventional HAZOP and AI-assisted knowledge capture: nodes, deviations, causes, consequences, safeguards); *detect* (digital twin monitoring and model-based early warning; is a known deviation, or its precursor, developing in real operation); *diagnose* (model-based or AI classification distinguishing between possible causes of the same deviation); and *respond* (connecting the diagnosed scenario to safeguard status, risk escalation, and recommended action). These four levels require different data, models, validation evidence, and governance controls, and should not be confused with one another. This logic also clarifies why digital twin and

AI technologies belong as decision-support layers rather than HAZOP replacements [29,56]: the HAZOP team remains essential for defining credible scenarios and judging consequence severity, while the technology layers help maintain that knowledge's relevance during operation by flagging changing assumptions, developing degradation, and approaching deviations, extending HAZOP's reach without removing the human expertise at its core. The remainder of this review examines each pathway in turn (AI-assisted HAZOP, digital twin architectures, hybrid physics–data models, and explainable AI), before assessing industrial maturity, implementation barriers, and the research agenda needed to move from fragmented demonstrations toward reliable HAZOP-informed digital twin systems.

3. AI-Assisted HAZOP: From Knowledge Automation to Expert Augmentation

Section 2 established that HAZOP knowledge has a structured form that can be encoded, searched, reused, and connected to operational data. This raises an important question: to what extent can artificial intelligence support or automate parts of the HAZOP workflow? The answer is more nuanced than a simple claim of automation. AI-assisted HAZOP has developed from early rule-based and expert-system approaches toward ontology-based models, knowledge graphs, natural language processing, and, more recently, large language models. These methods can support deviation generation, knowledge retrieval, consistency checking, documentation, and scenario suggestion. However, they do not remove the need for expert judgment, multidisciplinary review, or process-specific validation.

The reviewed literature indicates that AI-assisted HAZOP is best positioned as expert augmentation rather than full replacement. HAZOP is not only a mechanical exercise of applying guide words to process parameters; it is also a structured expert discussion that combines process knowledge, operating experience, equipment familiarity, human factors, and judgment about safeguard adequacy. AI can reduce repetitive effort, improve access to previous knowledge, and highlight omissions or inconsistencies, but it remains limited when asked to evaluate unusual causal pathways, judge safeguard independence, or decide whether recommendations are practical in a specific facility.

This section reviews the main AI-assisted HAZOP pathways. Section 3.1 explains why HAZOP automation has attracted sustained research interest. Section 3.2 reviews rule-based systems, ontology-based HAZOP, and knowledge graph approaches. Section 3.3 evaluates natural language processing and large language models. Section 3.4 compares the main AI-assisted HAZOP approaches. Finally, Section 3.5 provides a critical assessment of why AI-assisted HAZOP should be positioned as a decision-support layer within expert-led process safety practice.

3.1. Why HAZOP Automation Has Attracted Research Interest

Interest in AI-assisted HAZOP stems from a practical tension: HAZOP is valued for being systematic, auditable, and multidisciplinary [8,9], yet is time-consuming and resource-intensive, especially for large facilities with many nodes and recurring deviation patterns across comparable equipment [21–23]. Deviation generation, knowledge retrieval across similar equipment [57,58], cross-facility consistency checking, and documentation support are the parts of the workflow most suitable for computational assistance, directly addressing dependence on expert judgment, scalability, and reuse of previous-study lessons. However, the most safety-critical judgments, including whether a cause is credible in a specific context and whether a safeguard is independent and effective, depend on plant-specific design, operating history, and organizational procedures that should remain under qualified human authority. The literature is therefore best interpreted through an augmentation lens [56]:

a useful system helps the team work more systematically and retrieve knowledge more efficiently, not replace it, since overstating automation capability creates false confidence.

3.2. Rule-Based, Ontology-Based, and Knowledge Graph Approaches

Early computer-aided HAZOP relied on *rule-based and expert systems* encoding cause–deviation–consequence relationships explicitly (e.g., loss of cooling in an exothermic reactor leading to higher temperature and runaway risk). Early digraph-based and knowledge-based automation frameworks [32,58,59] offered transparency and auditability, but a subsequent state-of-the-art review found such approaches to be limited by the knowledge-engineering burden, poor transfer between process units, and difficulty in capturing non-linear dynamics or simultaneous deviations [29].

Ontology-based approaches formalize equipment types, parameters, deviations, causes, consequences, and safeguards as explicit, machine-readable relationships (e.g., linking heat-exchanger fouling to reduced heat transfer to a downstream high-temperature deviation), most substantially developed in a series of frameworks by Single and colleagues [17,18,58,60], building on earlier automated HAZOP-analysis tools such as PHASuite [61] and digraph-based models [32]. They support consistency, reuse, traceability, and linkage to digital twins, but require significant expert effort to build and may generate irrelevant scenarios if too generic, or become unmaintainable if too specific. *Knowledge graphs* extend this logic by linking equipment, deviations, incident histories, and MOC documents across sources, enabling retrieval of historical scenarios and consistency checking [19,62–64]. Their reliability depends on the coverage and accuracy of the underlying data, requiring expert review and terminology control during construction. Across all three approaches, the realistic role is structured knowledge support rather than autonomous hazard analysis, formalizing knowledge, reducing inconsistency, and improving retrieval, while output credibility still depends on expert validation.

3.3. Natural Language Processing and Large Language Models

NLP methods extract structured information from the unstructured text in which much process safety knowledge resides (HAZOP worksheets, incident reports, procedures, MOC records), supporting entity/relationship extraction, terminology standardization, and, combined with ontologies or knowledge graphs, populating structured knowledge bases from legacy documents [57,65]. Large language models have expanded interest further, generating fluent drafts, summaries, and candidate deviations [65–67]; recent evaluations of GPT-class models on STPA hazard identification [68], incident-based question-answering [69], and prompt-engineered HAZOP assessment [70] confirm that LLMs can support preliminary tasks but consistently require expert validation and hallucination control, since fluent text does not guarantee correct causal reasoning, adequate coverage, or physical consistency. This is a safety-governance issue, not merely a text-quality one.

The most defensible use of LLMs in HAZOP is therefore controlled and bounded. Acceptable roles (all subject to expert review): (1) document retrieval and preliminary HAZOP worksheet drafting for structured expert critique; (2) incident knowledge extraction and terminology standardization; (3) cross-referencing deviation scenarios against historical incident databases; (4) consistency checking across successive HAZOP revisions. Roles that must remain under competent human authority: (1) independent hazard identification without expert verification; (2) risk-ranking assignment; (3) safeguard adequacy judgment; (4) generation of final HAZOP recommendations. This conservative position reflects both the current validation evidence and the safety-critical stakes of HAZOP decisions; safety-critical LLM use also requires version control, prompt traceability, source attribution, and clear responsibility for final decisions [67,71–73].

3.4. Comparative Assessment of AI-Assisted HAZOP Approaches

The approaches reviewed above differ in their level of maturity, transparency, data requirements, and suitability for safety-critical use [29,56]. Table 7 summarizes their main contributions, strengths, limitations, and most defensible role in HAZOP enhancement. Supplementary Table S8 provides an expanded, source-referenced companion comparison, which additionally separates historical expert systems and adds a hybrid-approach category; it lists representative works, automation level, and TRL for each approach alongside a capability legend distinguishing industrially demonstrated, prototype-demonstrated, and not-yet-demonstrated capabilities.

Table 7. Comparison of AI-assisted HAZOP approaches and their roles in expert-led process safety practice.

Approach	Main Contribution	Key Strengths	Main Limitations and Appropriate Role
Rule-based and expert systems	Encode explicit cause–deviation–consequence logic	Transparent, auditable, and suitable for well-understood scenarios	Require extensive expert rule development; brittle outside encoded knowledge. Best used for structured deviation screening and transparent decision support.
Ontology-based HAZOP	Formalize HAZOP concepts and relationships	Supports consistent terminology, reasoning, traceability, and reuse	Requires major knowledge-engineering effort and expert validation. Best used as the semantic foundation for AI-assisted HAZOP and HAZOP-informed digital twins.
Knowledge graphs	Link HAZOP records, incidents, equipment, safeguards, and operational knowledge	Improves retrieval, organizational learning, and consistency checking	Depends on data quality and careful curation. Best used for knowledge management and scenario retrieval, not autonomous hazard judgment.
Natural language processing	Extract structured information from reports, worksheets, procedures, and incident narratives	Useful for legacy-document mining and terminology standardization	Extraction errors and contextual ambiguity require expert review. Best used to support knowledge-base construction and document search.
Large language models	Draft text, summarize documents, suggest candidate scenarios, and support review prompts	Flexible natural language interface and reduced documentation burden	Risk of hallucination, incomplete causal reasoning, and plausible but incorrect outputs. Best used only as a bounded drafting and retrieval assistant under expert supervision.

3.5. Critical Assessment: Why Full HAZOP Automation Remains Limited

The literature on AI-assisted HAZOP shows clear progress but persistent limitations [29,48]: HAZOP reasoning is inherently contextual (deviation significance, cause credibility, and safeguard adequacy all depend on process-specific and situational judgment that should remain under qualified human authority); validation evidence remains largely limited to simplified case studies and retrospective document analysis rather than diverse industrial facilities and abnormal operating modes, with false omissions in particular under-evaluated relative to correct suggestions; and governance requirements (traceability of data, model version, assumptions, and expert approval) are not yet consistently met, risking automation bias and unclear accountability if left unaddressed. The near-term future of AI-assisted HAZOP is therefore expert augmentation rather than full automation, consistent with the lifecycle safety-intelligence framework in Section 2: AI strengthens the reach and usability of HAZOP knowledge without replacing the engineering judgment that process safety depends on.

The findings of this section also motivate the next technology pathway. If AI-assisted HAZOP can help formalize and organize safety knowledge, the next question is how that knowledge can be connected to operational data so that known deviations can be monitored during plant operation. Section 4 therefore examines digital twin architectures for dynamic hazard monitoring and early warning.

4. Digital Twin Architectures for Dynamic Hazard Monitoring

Section 3 examined how AI-assisted HAZOP can support knowledge capture, scenario retrieval, documentation, and expert augmentation. However, formalizing HAZOP knowledge is only one part of the transformation. For HAZOP to become more operationally useful during the plant lifecycle, its deviation–cause–consequence–safeguard logic can be connected to live process data, process models, equipment condition, and operator decision support. This is the role of digital twin architectures in DT–AI-enhanced HAZOP.

In this review, a HAZOP-informed digital twin is understood as a safety-support architecture that links structured HAZOP knowledge with live operational data and model-based reasoning. Its purpose is not to replace alarms, safety instrumented systems, process control systems, or expert judgment. Rather, it provides an additional advisory layer that helps identify whether known HAZOP deviations are developing, whether process conditions are moving closer to hazardous limits, whether safeguards should be checked, and whether operators require earlier contextual warning. This section therefore examines how digital twin architectures can address one of the central limitations of conventional HAZOP: the separation between static hazard documentation and dynamic plant behavior.

4.1. From Process Monitoring to HAZOP-Informed Digital Twins

Conventional process monitoring treats abnormal conditions as variable excursions against alarm limits rather than structured hazard scenarios [28]: a high-temperature alarm indicates a limit was exceeded, but does not explain which HAZOP deviation is developing, which causes are plausible, or which safeguards require verification. A HAZOP-informed digital twin extends this by interpreting live measurements in relation to known deviations, causes, consequences, and safeguards [15,16,38,39,45]; for example, interpreting gradual heat-transfer loss as a precursor to a high-temperature deviation, or rising column pressure drop as a precursor to flooding. The conceptual shift is from “is the variable outside its limit?” to “is the process moving toward a known hazardous deviation, what causes are likely, and what safeguards should be verified?” This marks a shift from variable monitoring to scenario-aware monitoring.

4.2. Core Architecture of a HAZOP-Informed Digital Twin

A process safety digital twin can be represented as a five-layer architecture [28]. The *physical asset layer* includes equipment, instrumentation, alarms, interlocks, safety instrumented functions, relief devices, and human operators, and generates the process data needed for monitoring and diagnosis. The *data acquisition and integration layer* collects data from DCS, historians, laboratory and maintenance systems, and alarm databases; its reliability, sampling frequency, calibration status, and contextual meaning directly determine the credibility of downstream outputs. The *model layer* comprises first-principles, reduced-order, empirical, machine learning, or hybrid physics–data models that are used to estimate hidden states and forecast trajectories before a conventional alarm threshold is crossed [15,38,46,51,52].

The fourth layer, the *structured HAZOP knowledge layer*, is the feature that distinguishes a HAZOP-informed digital twin from a generic plant digital twin. This layer contains nodes, parameters, deviations, causes, consequences, safeguards, and operating assumptions; without it, the digital twin may detect anomalies but cannot explain their safety significance. Finally, the *analytics and decision-support layer* converts data, model outputs, and HAZOP knowledge into deviation recognition, fault diagnosis, time-to-limit estimation, safeguard-status checking, and explainable operator advisory support.

Figure 3 illustrates the proposed layered architecture of a HAZOP-informed digital twin for dynamic hazard monitoring.

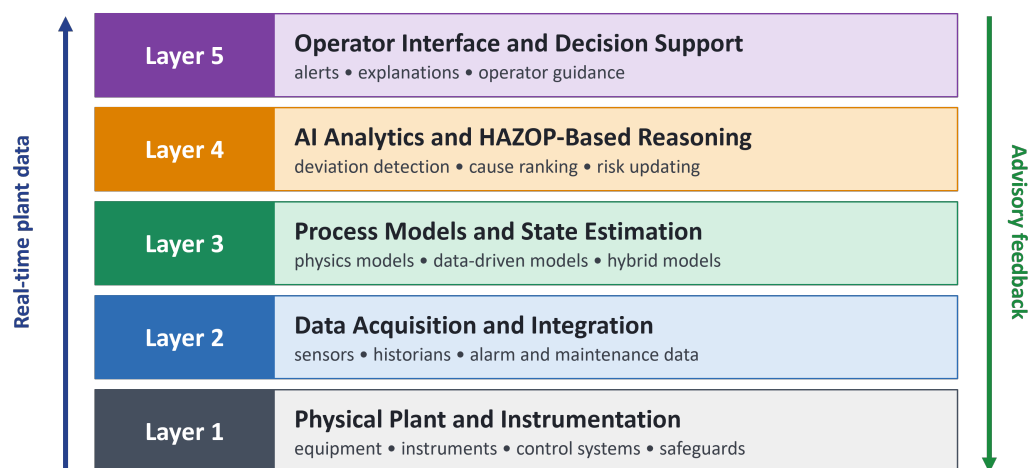


Figure 3. Layered architecture of a HAZOP-informed digital twin for dynamic hazard monitoring. The architecture links physical plant data, data acquisition, process models, structured HAZOP knowledge, AI analytics, and operator decision support to transform static HAZOP knowledge into lifecycle safety intelligence.

4.3. Dynamic Deviation Monitoring and Early Warning

One important function of a HAZOP-informed digital twin is detecting movement toward documented deviations before they become severe enough to trigger a conventional alarm, as fouling, catalyst deactivation, sensor drift, or feed changes accumulate [74,75]. This requires mapping deviations to measurable indicators (e.g., “high pressure” to a direct measurement) or inferred ones (e.g., reduced heat-transfer capability to an estimated overall heat-transfer coefficient), and forecasting trajectories rather than only the current state, so that gradually developing margin loss (a slow reactor temperature rise, a progressive cooling-duty reduction) is connected to HAZOP deviation logic before it violates a fixed alarm limit [38,46,47]. Threshold design requires care: making them too sensitive risks alarm fatigue, while making them too conservative fails to provide a useful warning, so validation against operating envelopes and known incident mechanisms is essential.

4.4. Fault Diagnosis and Cause Ranking

A single deviation may have several possible causes (e.g., high reactor temperature from cooling failure, excessive feed concentration, agitator malfunction, or reaction-rate acceleration) [75,76], and a HAZOP-informed digital twin can support diagnosis by integrating process measurements, model residuals, equipment-health indicators, and HAZOP cause–consequence logic to indicate which causes are most consistent with the evidence [34,46,47]. This is valuable because different causes require different responses (checking cooling-water supply versus adjusting feed composition). This output must be presented with uncertainty and supporting evidence rather than as a final diagnosis, since diagnostic models can be misleading when the plant operates outside validated conditions, sensors fault, or multiple faults occur simultaneously.

4.5. Safeguard Monitoring and Protection-Layer Awareness

Linking HAZOP knowledge to a digital twin architecture [28] also enables monitoring whether documented safeguards (alarms, interlocks, relief devices, independent protection layers) remain available, tested, bypassed, or effective during operation, consistent with hazard evaluation and risk-based process safety practice [8,35]; for example, checking whether a level transmitter is healthy and an alarm is active rather than inhibited, or whether a safety instrumented function is under maintenance or bypass. This matters because many major accidents involve safeguards that existed on paper but were unavailable

or ineffective under actual operating conditions. A digital twin cannot replace formal safety lifecycle management, but it can provide a current view of whether HAZOP assumptions remain valid and support more meaningful prioritization: a developing deviation with healthy safeguards may require monitoring, while the same deviation with degraded safeguards may require urgent action.

4.6. Functional Summary of HAZOP-Informed Digital Twins

The preceding subsections show that HAZOP-informed digital twins can support several distinct but connected safety functions. Table 8 summarizes these functions and illustrates how each contributes to dynamic hazard monitoring. The table also reinforces an important point: the value of a HAZOP-informed digital twin is not only its ability to detect abnormal data patterns, but its ability to translate those patterns into HAZOP-relevant safety information that supports operator and engineering judgment.

Table 8. Key functions of HAZOP-informed digital twins for dynamic hazard monitoring.

Function	Purpose	Example HAZOP-Related Output
Dynamic deviation monitoring	Detect movement toward known HAZOP deviations using live process data and model estimates.	Early warning that reactor temperature is approaching a documented “MORE temperature” deviation.
Cause ranking	Identify which documented causes are most consistent with current evidence.	Cooling failure ranked above feed-concentration disturbance as the likely cause of high reactor temperature.
Safeguard-status monitoring	Indicate the availability, inhibition, maintenance, diagnostic, or demand status of documented safeguards where suitable data are available.	Warning that a relevant alarm is inhibited, an interlock is bypassed, or a safety function is under maintenance; engineering verification remains required.
Time-to-limit prediction	Estimate the time remaining before an operating or safety limit is reached.	Estimated time before column pressure drop reaches a flooding-related limit.
Operator advisory support	Present scenario context, supporting evidence, relevant safeguards, and recommended checks.	Display of likely HAZOP scenario, supporting trends, and safeguards to verify.

4.7. Implementation Constraints and Data Requirements

Despite its potential, a HAZOP-informed digital twin must address five practical implementation constraints [28]: *data quality* (reliable sensors, correct tags, validated historian data, and clear links to HAZOP nodes, because poor quality causes false alerts or missed deviations); *model validity* (a model performing well under normal conditions may fail under abnormal, transient, or degraded-equipment scenarios, requiring explicit validation boundaries and out-of-domain warnings); *integration* with existing management-of-change, alarm-management, maintenance, and incident-learning systems, without which the digital twin becomes an isolated tool; *human factors*, since too many alerts or overly authoritative outputs erode trust or encourage automation bias [49,77], making interface and governance design as important as the model itself; and *cybersecurity and data governance*, since a digital twin connected to live plant data and safety logic must be protected against unauthorized access, tampering, and availability loss through access control, audit trails, and version management.

4.8. Critical Assessment of Digital Twin-Based HAZOP Enhancement

Digital twin architectures offer an important pathway for enhancing HAZOP because they address the gap between static hazard review and changing operational re-

ality that conventional HAZOP cannot easily solve [28], but should not be overstated: many current applications remain conceptual, simulation-based, or limited to specific equipment [15,16,38,46], with fewer studies demonstrating fully integrated systems connecting live plant data to structured HAZOP knowledge, validated diagnostics, safeguard-status monitoring, and operator decision support. The near-term role supported by the reviewed literature is as an advisory safety layer that helps relate current process behavior to known HAZOP scenarios, with explainable, validated, and auditable outputs, not as a replacement for alarms, safety instrumented systems, or expert judgment.

The findings of this section motivate the next pathway: hybrid physics–data models. A digital twin can provide the architecture for dynamic monitoring, but its predictive value depends heavily on the quality of the models used to estimate process behavior, forecast deviations, and support diagnosis. The next section therefore examines how hybrid physics–data modeling can strengthen predictive safety analytics in DT–AI-enhanced HAZOP.

5. Hybrid Physics–Data Models for Predictive Safety Analytics

Section 4 examined the digital twin architecture used to connect structured HAZOP knowledge with live process data, model-based monitoring, and operator decision support. However, the value of a HAZOP-informed digital twin depends strongly on the quality of the models used inside it. Meaningful early warning requires estimating the current process state, predicting how abnormal conditions may evolve, and relating those predictions to known HAZOP deviations. This is where hybrid physics–data models become important.

Hybrid physics–data models combine first-principles process knowledge with data-driven learning. In process safety, this combination is attractive because purely mechanistic and purely data-driven models each have limitations. First-principles models provide physical consistency and extrapolation discipline, but they may be incomplete, expensive to maintain, or too slow for real-time use. Data-driven models can learn complex patterns from plant data, but they may fail when the process moves outside the training domain, when sensors drift, or when rare abnormal events are not well represented in historical data. Hybrid models attempt to balance these limitations by combining physical structure with statistical adaptability.

In the context of HAZOP enhancement, the role of hybrid modeling is not simply to improve prediction accuracy. Its deeper contribution is to support predictive safety reasoning. A hybrid model can estimate whether a process is moving toward a known deviation, forecast the time available before a safety or operating limit is reached, distinguish between alternative causes of the same deviation, and quantify uncertainty in the prediction. These functions can help transform HAZOP from a static record of possible deviations into a dynamic source of predictive safety intelligence.

5.1. Why Predictive Models Matter for HAZOP Enhancement

Conventional HAZOP identifies credible deviations, causes, consequences, and safeguards but does not continuously predict when a deviation may develop during operation, even though many hazardous scenarios evolve gradually, including fouling, catalyst deactivation, compressor degradation, and sensor drift, reducing the safety margin assumed during the original study without immediately triggering an alarm [74,78]. Predictive models estimate process behavior before a conventional alarm limit is crossed (e.g., forecasting that reactor temperature will exceed a safe limit as cooling deteriorates, or that a column is approaching flooding based on pressure-drop trends), and this role is closely connected to HAZOP logic: a prediction is most useful when interpreted through the node–deviation–cause–consequence–safeguard framework rather than as a bare statistical forecast. Predictive modeling for process safety is more demanding than ordinary process

optimization, requiring credibility under abnormal and transient conditions and honest uncertainty communication; prediction should be treated as an advisory input to expert judgment, not an automatic decision-maker.

5.2. First-Principles, Data-Driven, and Hybrid Modeling Approaches

Three broad modeling approaches are relevant to predictive safety analytics, each with limitations that become important in safety-critical use. *First-principles models* describe process behavior using conservation laws, thermodynamics, reaction kinetics, and equipment-performance relationships (e.g., a reactor energy balance linking heat generation, heat removal, and cooling performance); their physical grounding makes outputs interpretable and linkable to HAZOP deviations, but parameters such as heat-transfer coefficients, catalyst activity, or fouling resistance drift over time, and detailed dynamic models can be costly to build, calibrate, and maintain. *Data-driven models* instead learn relationships directly from historical or real-time process data using statistical process monitoring, machine learning, or sequence models [46,66]; they are effective where instrumentation and historical data are abundant, but may learn correlations without capturing the underlying physical mechanism, and can appear accurate on normal data while failing during rare abnormal events, start-up, shutdown, or conditions outside the training domain, which are precisely the conditions that matter most for HAZOP-informed early warning.

Hybrid physics–data models combine both approaches: a first-principles baseline with a data-driven residual correction; physical constraints imposed on a learned model; or physics-informed machine learning, in which governing equations are embedded directly into training [51,79–81]. This combination is well suited to HAZOP-informed digital twins when properly validated: the physical component helps keep predictions physically realistic and engineering-interpretable, while the data-driven component adapts to plant-specific behavior and degradation, together supporting prediction, diagnosis, and uncertainty-aware decision support with more interpretability than a purely black-box model.

Figure 4 illustrates how first-principles process knowledge and plant data can be combined in hybrid physics–data models to support predictive safety analytics and HAZOP-informed decision support.

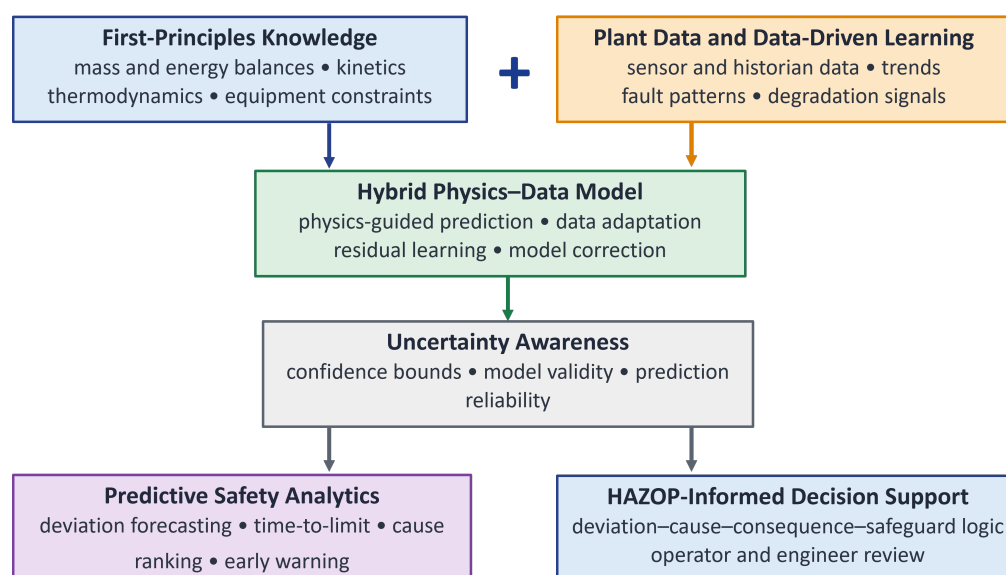


Figure 4. Hybrid physics–data modeling framework for predictive safety analytics in HAZOP-informed digital twins. First-principles process knowledge and plant data are combined to support deviation forecasting, time-to-limit prediction, cause ranking, uncertainty awareness, and HAZOP-informed decision support before conventional alarm limits are reached.

5.3. Model-Based Early Warning and Time-to-Limit Prediction

Conventional alarms activate only when a variable crosses a predefined threshold, providing limited warning for slowly developing conditions; hybrid models extend this by estimating how much time remains before an operating or safety limit is reached [47,74,75,78] (e.g., estimating remaining time before a reactor approaches a high-temperature limit as cooling duty decreases, or before column flooding risk becomes significant). This has direct HAZOP value: a predicted trajectory can be translated into HAZOP language (moving toward “MORE temperature”, “LESS flow”) and connected to likely causes, consequences, and safeguards, making early warning a process safety interpretation rather than a bare statistical alert, provided the estimate’s confidence is communicated alongside the time-to-limit itself, and thresholds are validated against expert-reviewed abnormal scenarios to avoid both alarm fatigue and missed escalation.

5.4. Fault Diagnosis and Deviation–Cause Discrimination

A single deviation may have several possible causes (a high-temperature deviation from cooling failure, excessive reactant concentration, fouling, or agitator malfunction) [47,75,76], and hybrid models support discrimination by comparing measured behavior against physics-based expected behavior, with data-driven components classifying the resulting residual patterns to rank likely causes [75,76]. The HAZOP worksheet’s documented cause list provides a structured diagnostic frame, so the model evaluates which known causes are most consistent with current evidence rather than searching an unlimited fault space, but this must be presented as ranked evidence, not a final conclusion, since multiple faults can co-occur, sensor faults can mimic process faults, and operation can move outside the model’s validated region.

5.5. Uncertainty, Model Drift, and Validation

Because an incorrect prediction carries real safety consequences (a false negative missing a developing hazard; a false positive contributing to alarm fatigue), hybrid models should communicate confidence and validity alongside safety-relevant predictions [78,82]. This includes measurement uncertainty (sensor noise, drift), parameter uncertainty (imprecisely known heat-transfer coefficients, fouling factors), structural uncertainty (model–plant mismatch), and operational uncertainty (feed, rate, or equipment changes).

Model drift is a safety issue, not only a performance issue, since a stale model may understate or overstate the margin to a hazardous deviation as equipment ages or operating strategy changes. This requires residual monitoring, periodic recalibration, and clear model-management procedures. Validation should also extend beyond historical-data fitting to abnormal scenarios, transients, and known incident mechanisms, with particular attention to avoiding dangerous omissions, since missing a credible hazardous trajectory is more serious than flagging extra cases for review; established ML safety-assurance principles for lifecycle validation and operational-boundary treatment apply directly here [82].

5.6. Comparative Assessment of Modeling Approaches

Table 9 summarizes the main modeling approaches relevant to predictive safety analytics and their roles in HAZOP-informed digital twins [47,76,78]. Supplementary Table S2 provides a source-referenced comparison of standard PINNs, physics-informed recurrent modeling, and an industrial hybrid soft-sensor application, including demonstrated application, reported performance, TRL, and key limitations.

Table 9. Comparison of modeling approaches for predictive safety analytics in HAZOP-informed digital twins.

Modeling Approach	Main Contribution	Key Strengths	Limitations and Suitable Role
First-principles models	Represent process behavior using conservation laws, thermodynamics, kinetics, and equipment equations.	Physically meaningful; useful for scenario analysis and engineering interpretation.	Require uncertain parameters and may be difficult to calibrate or maintain in real time. Best used as the physical backbone for safety prediction and scenario testing.
Data-driven models	Learn operating patterns from plant measurements and historical data.	Useful for anomaly detection, classification, forecasting, and multivariable pattern recognition.	May extrapolate poorly outside training data and may lack causal meaning. Best used for advisory monitoring with clear validation boundaries and expert interpretation.
Hybrid physics–data models	Combine physical structure with data-driven adaptation, residual learning, or model correction.	Balance physical consistency with plant-specific learning and improved predictive capability.	Require careful integration, calibration, uncertainty management, and model governance. Best used for predictive safety monitoring, deviation forecasting, and time-to-limit estimation.
Physics-informed machine learning	Embed physical equations, constraints, or residuals into the learning process.	Can improve physical plausibility, generalization, and consistency with known process behavior.	Can be complex to implement for large industrial systems and still requires validation. Best used when governing equations are known but states or parameters are uncertain.
Probabilistic and Bayesian models	Update beliefs about causes, deviations, or safeguard failures as new evidence becomes available.	Useful for uncertainty-aware diagnosis, cause ranking, and dynamic risk updating.	Require well-defined conditional relationships and careful interpretation. Best used for cause ranking, uncertainty communication, and dynamic risk support.

5.7. Critical Assessment of Hybrid Models for HAZOP Enhancement

Hybrid physics–data models offer an important pathway for HAZOP enhancement, addressing static hazard review’s limited ability to predict how safety margins change during operation [74,76,78], but their role should not be overstated: a hybrid model is not automatically safe because it contains physics nor automatically accurate because it uses plant data, and it should be validated, monitored, and explicitly connected to HAZOP deviation–cause–consequence–safeguard logic rather than left as a disconnected predictive tool. The reviewed literature supports their role as predictive decision support, helping assess whether a known deviation is developing and how quickly a hazardous limit may be approached, with final judgment on risk significance and required action remaining with qualified personnel.

The findings of this section motivate the next pathway: explainable AI. Predictive models are more useful for safety decisions when their outputs are understandable, trustworthy, and auditable. The next section therefore examines explainable AI, operator trust, and regulatory acceptance in DT–AI-enhanced HAZOP.

6. Explainable AI, Operator Trust, and Regulatory Acceptance

Section 5 examined how hybrid physics–data models can support predictive safety analytics by forecasting deviations, estimating time-to-limit, ranking possible causes, and communicating uncertainty. However, prediction alone is not sufficient for process safety. In high-hazard operations, an AI-generated alert or recommendation should be understandable, technically credible, auditable, and acceptable to the people who must act on it.

This makes explainable AI (XAI), operator trust, and regulatory acceptance important to DT–AI-enhanced HAZOP.

Explainable AI is especially important because HAZOP is built on transparent expert reasoning. A conventional HAZOP worksheet records the logic connecting a deviation to its causes, consequences, safeguards, and recommendations. If AI or digital twin systems are introduced into this workflow, their outputs should preserve this traceability. A black-box model that predicts a hazardous state without explaining the supporting evidence may have limited practical value, even if its statistical performance appears strong. Operators, process engineers, safety specialists, and regulators need to understand why the system produced a warning, what assumptions were used, which variables contributed to the result, and whether the recommendation lies within the validated operating domain of the model [82–84].

The reviewed literature indicates that explainability should not be treated as an optional visualization feature added after model development. For HAZOP-informed digital twins, explainability should be built into the safety-support architecture. Explanations should be expressed in process safety terms: deviation, cause, consequence, safeguard, uncertainty, and recommended verification. This section reviews the role of XAI in safety-critical decision support, the relationship between explanation and operator trust, the regulatory and auditability requirements for DT–AI-enhanced HAZOP, and the limitations of explainability as a risk-control measure.

6.1. Why Explainability Matters in HAZOP-Informed Digital Twins

In process safety, explanation carries a stronger requirement than in general industrial AI: a safety-support system that cannot explain its reasoning is difficult to trust, validate, or defend during audits or incident investigations, since decisions may involve hazardous materials, high-energy systems, and risks to personnel. A useful explanation should connect the output to HAZOP reasoning, not merely identify influential input variables. For example, a digital twin’s reactor warning should indicate the relevant deviation (“MORE temperature”), likely causes (reduced cooling duty, increased reaction rate), possible consequences, and safeguards to verify, which is more useful than a generic feature-importance plot. This matters because common XAI techniques (feature attribution, surrogate models, counterfactuals, SHAP [85], LIME [86]) were developed for general machine learning and do not automatically produce safety-relevant explanations without engineering translation into HAZOP terms [83,84]. Explainability also supports accountability: the system should record which model version was used, what data and assumptions applied, and what human action followed, preserving the same auditable discipline as HAZOP records and MOC documentation.

6.2. Interpretability Versus Explainability: A Critical Distinction for Safety

A distinction often blurred in the AI literature but consequential for safety-critical applications is *interpretability* (a model whose internal structure, such as decision trees, linear models, or Bayesian networks, can be directly examined by a qualified human) versus *explainability* (an internally opaque model, such as a deep neural network or LLM, accompanied by a post hoc approximation of why an output was produced) [25,85,86]. Interpretable models offer a safety advantage: their reasoning can be reviewed by engineers, auditors, and regulators without additional tools, closely aligned with process safety’s documentation culture, whereas post hoc explanations may not fully capture the true decision boundary, may be unstable across similar inputs, and can be locally plausible but globally misleading [87]. The practical implication is that interpretability should be preferred wherever model performance is adequate; where complex models are needed,

post hoc explanations should be used with explicit acknowledgment of their limitations and not treated as verified causal accounts.

6.3. Forms of Explainability Relevant to Process Safety

Five forms of explainability are relevant to DT–AI-enhanced HAZOP. *Model transparency*: Rule-based systems, decision trees, linear models, Bayesian networks, and simplified mechanistic models can be inspected more directly than deep neural networks, but simple models may not capture complex nonlinear process behavior. *Post hoc explanation*: Methods such as SHAP [85] and LIME [86] explain a complex model's outputs after prediction and are among the most widely applied in process safety fault detection and anomaly classification [83,84,87], but a feature-importance result showing that temperature, flow, and concentration influenced a prediction does not by itself confirm the causal mechanism or safeguard adequacy, so engineering interpretation remains necessary. *Physics-grounded interpretation*: In hybrid physics–data models, explanations grounded in mass/energy balances, thermodynamic constraints, or equipment-performance indicators (e.g., a decreasing heat-removal margin or rising fouling resistance) connect model outputs to familiar engineering concepts. Supplementary Table S3 compares six methods spanning these forms (SHAP, LIME, counterfactual explanations, decision trees/rule-based methods, physics-grounded interpretation, and ontology-based reasoning), with their explanation output, strengths, limitations, representative process safety application, and TRL.

The two forms most important for this review are *HAZOP-structured explanation*, in which the AI system's internal reasoning (machine learning, probabilistic, or hybrid) is translated into the node–parameter–deviation–cause–consequence–safeguard–recommendation structure process safety teams already use, making the system easier to review and more compatible with existing safety-management practice; and *uncertainty explanation*, in which the system communicates not only its prediction but its confidence and the source of any uncertainty (sensor quality, missing data, model extrapolation, conflicting evidence), so that a high-uncertainty warning is presented differently from one supported by strong evidence.

6.4. Operator Trust, Human Factors, and Regulatory Governance

Trust is important for adoption: distrust causes operators to ignore warnings, while excessive trust causes uncritical acceptance; both are unsafe. The aim is *calibrated* trust, based on technical reliability, explanation quality, consistency with process knowledge, interface design, and clear governance of the system's advisory status [88,89]. Human-in-the-loop design should therefore support recognition, diagnosis, prioritization, and response while final judgment remains with qualified personnel, especially for shutdown, bypass, or emergency decisions.

Two human-factor risks require explicit control. *Automation complacency* occurs when operators reduce independent verification because the system is usually correct, a risk well documented in the situation-awareness and calibrated-trust literature [88–90]; mitigation includes uncertainty communication, explanations grounded in HAZOP terminology, graceful degradation outside the validated envelope, and moderate rather than high automation levels [83,84,87]. *Skill atrophy* occurs when routine AI support weakens the process knowledge needed for novel or edge-case scenarios; mitigation includes human-led review of complex scenarios, independent verification exercises, and training that uses AI as a support tool rather than a substitute for expertise [49,56,88,91]. Safety management systems should also define who verifies AI outputs, how conflicts with expert judgment are resolved, and how system performance is monitored over time [56,88,91–93].

Regulatory acceptance depends on auditability, model governance, and a clear boundary between advisory support and safety-critical control, consistent with established process safety management standards [8,9,36,37] and regulatory frameworks such as the OSHA Process Safety Management standard [21]. Evidence maturity should be stated honestly: a simulation-demonstrated method is not equivalent to an industrially validated tool, particularly for generative AI, where plausible outputs may conceal missing scenarios [25]. Expectations derived from emerging frameworks such as the EU AI Act, ISO/IEC 42001, and IEC 61508/UL 4600 (Section 8.14) do not yet address HAZOP tools specifically, but they provide useful governance expectations for safety-critical AI [62,94].

6.5. Failure Modes of Explainable AI in Safety-Critical Use

Although explainability is important, it is not sufficient [56]. Explanations can fail in several ways. First, they may be superficial: an alert may display “SHAP value for temperature = 0.8” without linking it to a specific “MORE temperature” deviation, operating threshold, consequence, or safeguard, creating false confidence without actionable process context (e.g., a feed-preheater warning above a site-defined 340 °C limit should identify the runaway consequence and cooling safeguard). Second, they may be technically correct but operationally unhelpful: a feature-importance plot may identify influential variables without telling the operator which safeguard to check or which procedure to follow (e.g., a reactor overpressure explanation should direct verification of cooling-water flow, relief availability, and feed isolation).

Third, explanations may imply causal meaning where only statistical association exists. For example, a model trained on normal distillation-column operation may provide high-confidence SHAP attributions during startup, even though startup was outside the training domain and the explanation has no validated causal basis. Fourth, explanations may overload the user or become unstable: repeated LIME explanations for the same event can alternate between cooling-water flow and reactor pressure as the dominant contributor, which is problematic during abnormal operation. Fifth, clear explanations can still encourage overtrust if they omit uncertainty, validity limits, or out-of-domain warnings (e.g., repeated “within-bounds” messages while ammonia synthesis loop pressure trends upward across several shifts). For this reason, XAI outputs should be layered, concise for operators and deeper for engineering review, and linked to model-validity status.

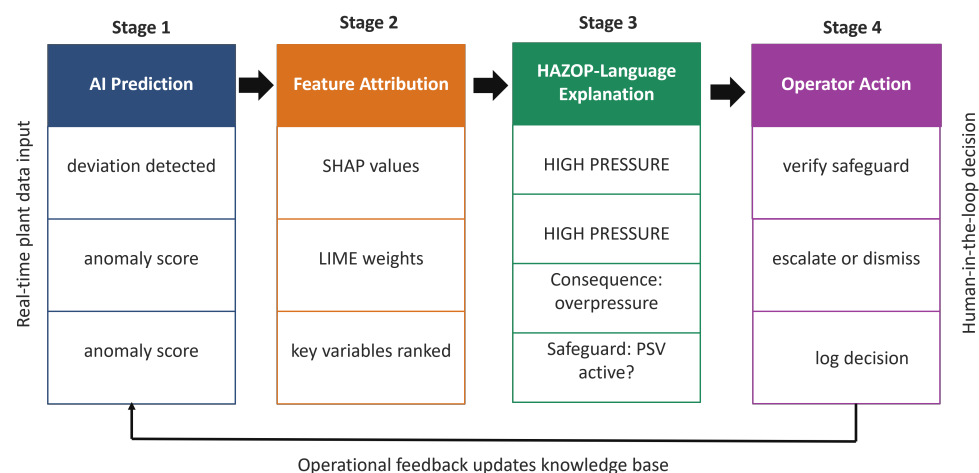
These failure modes show that XAI should be treated as part of a broader risk-control strategy. It should be combined with model validation, human factors design, governance, training, and clear accountability. In this sense, XAI is not a substitute for safety assurance; it is one element of a wider assurance framework for AI-supported process safety [82].

6.6. Practical Requirements for Explainable DT-AI HAZOP Support

For DT-AI-enhanced HAZOP to be credible in industrial practice, explainability should be designed around seven practical requirements [56,88,91]: HAZOP-aligned language, evidence-based reasoning, uncertainty awareness, role-specific detail, auditability, human-in-the-loop control, and lifecycle-learning feedback. Table 10 summarizes each requirement, its purpose, and its practical implementation implication. Figure 5 illustrates the corresponding explainability chain, in which real-time plant data are converted into AI predictions, feature-attribution evidence, HAZOP-language explanations, and operator-action options, with feedback returned to the knowledge base for lifecycle learning.

Table 10. Practical explainability requirements for HAZOP-informed digital twins.

Requirement	Purpose in DT-AI HAZOP Support	Practical Implementation Implication
HAZOP-aligned explanation	Express AI outputs using process safety language rather than generic model terms.	Alerts should identify the relevant node, deviation, likely causes, possible consequences, safeguards, uncertainty, and recommended verification.
Evidence-based reasoning	Allow operators and engineers to understand why an alert or recommendation was generated.	The interface should show supporting measurements, model residuals, process trends, historical cases, and relevant HAZOP records.
Uncertainty awareness	Prevent overconfidence in model outputs and support cautious decision making.	The system should report confidence level, model validity status, missing data, sensor-quality concerns, and whether the process is outside the validated model domain.
Role-specific explanation	Provide the right level of detail for different users.	Operators may need concise action-focused explanations, while process engineers, safety specialists, and auditors may need deeper diagnostic and historical evidence.
Auditability and traceability	Support regulatory review, incident investigation, and internal safety governance.	The system should preserve records of input data, model version, generated alert, explanation shown, human decision, and follow-up action.
Human-in-the-loop control	Maintain expert judgment and accountability in safety-critical decisions.	AI outputs should remain advisory unless formally validated and governed; final judgment should remain with qualified personnel and approved procedures.
Lifecycle learning	Ensure that operational experience improves future hazard analysis and safety management.	Confirmed alerts, rejected alerts, near misses, model drift, and new causal pathways should feed back into HAZOP revalidation, MOC, alarm management, procedures, and training.

**Figure 5.** XAI explainability chain for HAZOP-informed digital twin support. Real-time plant data are converted into AI predictions, feature-attribution evidence, HAZOP-language explanations, and operator-action options, with feedback returned to the knowledge base for lifecycle learning.

6.7. Critical Assessment of XAI for HAZOP Enhancement

Explainable AI provides an important pathway for DT-AI-enhanced HAZOP because it addresses the human and organizational side of digital safety support [88,91], but should not be oversold: an explanation does not prove a model is correct, that training data were representative, or that a recommendation generalizes to abnormal conditions; explainability is an important but not sufficient condition for safety. XAI is best positioned to make HAZOP-informed advisory outputs understandable, reviewable, and auditable (which deviation, which plausible causes, which safeguards to verify, how reliable the output is), with the final decision remaining within expert-led process safety practice.

The findings of this section complete the four technology pathways examined in Sections 3–6. AI-assisted HAZOP supports knowledge capture and reuse; digital twin architectures connect hazard knowledge to operational data; hybrid physics–data models enable predictive safety analytics; and explainable AI supports trust, auditability, and regulatory acceptance. The next section therefore evaluates the industrial maturity of these pathways and the extent to which they have moved from conceptual or prototype demonstrations toward practical deployment.

7. Industrial Applications and Technology Maturity

Sections 3–6 examined four complementary pathways for DT–AI-enhanced HAZOP: AI-assisted HAZOP, digital twin-based monitoring, hybrid physics–data models, and explainable AI. These pathways show clear conceptual potential, but their industrial maturity is uneven. Some methods have been demonstrated through case studies, simulations, prototypes, or equipment-specific applications, while fully integrated HAZOP-informed digital twin systems remain limited.

This section evaluates the maturity of the four pathways from an implementation perspective. The aim is not simply to list applications, but to distinguish between conceptual proposals, laboratory or simulation demonstrations, prototype systems, and industrially validated tools. This distinction is important because process safety adoption requires more than technical feasibility. A method also needs evidence of reliability under abnormal conditions, explainability to users, compatibility with existing safety-management systems, auditability, maintainability, and acceptance by plant personnel and regulators.

7.1. Evidence Maturity Across the Four Technology Pathways

The four pathways differ significantly in their current level of evidence. AI-assisted HAZOP has a relatively long research history, especially in rule-based, ontology-based, and computer-aided HAZOP methods. These studies show that parts of HAZOP knowledge can be formalized and reused computationally [32,35]. Ontology-based methods and knowledge graphs have extended this formalization [18,19]. Machine learning and NLP approaches have extended this to automated deviation generation and risk classification [60,61]. More recent work on process safety knowledge graphs and NLP-based extraction shows potential for knowledge retrieval [19,65]. LLM-assisted HAZOP approaches have demonstrated capability for documentation support and scenario generation [68,69]. Additional applications of LLM to HAZOP and asset integrity have been demonstrated [70,72]. However, most of these applications remain closer to expert-support prototypes than fully validated industrial HAZOP systems.

Digital twin applications in process safety show increasing maturity, particularly for monitoring, visualization, operator support, and equipment-level safety applications. Digital twin concepts have been applied to operator-risk prevention in process plants, petrochemical safety improvement, hydrogen refueling station safety, and broader safety-risk analysis [15,16]. Process safety digital twin implementations have demonstrated monitoring and decision-support capabilities [38,46]. More recent applications include process safety management in oil and gas operating units [40], cloud-integrated digital twin architectures for real-time risk monitoring in oil and gas operations [41], dynamic PSM architectures integrating digital twins with machine learning [42], and safety digital twins for industrial collaborative robots [44]. These studies demonstrate that digital twins can support safety monitoring and decision making. Nevertheless, many applications remain focused on general monitoring, simulation, visualization, or equipment health, rather than explicit integration with structured HAZOP deviation–cause–consequence–safeguard logic.

Hybrid physics–data models are technically promising because they support prediction under complex and uncertain process conditions. Process fault detection and diagnosis provide well-established methodological foundations for predictive safety analytics, as surveyed in [95,96]. Physics-informed learning methods have demonstrated capability for process monitoring and fault detection [79–81]. Hybrid physics–data models have been applied to predictive safety analytics in industrial process settings [53,54,97]. However, the industrial maturity depends on the specific application. Some model-based monitoring and fault diagnosis methods are well established in process systems engineering, while integrated hybrid models for HAZOP-informed early warning remain less mature. Their use in safety-critical decision support still requires additional validation under abnormal, transient, degraded, and rare-event conditions.

Explainable AI is important for operator trust and regulatory acceptance, but its maturity in process safety-specific applications is still developing. General XAI methods and frameworks are well documented [83,84], and process safety-oriented XAI approaches have been developed [87,98], and safety assurance for machine learning-based systems has been discussed in broader engineering contexts [82]. However, translating XAI outputs into HAZOP-specific explanations remains an unresolved need. Process safety users do not only need feature importance or model transparency; they need explanations expressed in terms of deviation, cause, consequence, safeguard, uncertainty, and recommended verification.

7.2. Technology Readiness Interpretation

The technology readiness level (TRL) is a useful, although imperfect, way to discuss a technology's maturity [29,56]. In the context of DT–AI-enhanced HAZOP, the lowest maturity level corresponds to conceptual frameworks or theoretical proposals. Intermediate maturity corresponds to simulation studies, retrospective case studies, proof-of-concept tools, or prototype systems tested on limited process examples. Higher maturity would require sustained industrial deployment, validated performance under real operating conditions, integration with existing safety-management systems, documented governance, and evidence of user acceptance.

Most DT–AI approaches for HAZOP enhancement currently appear to sit in the intermediate range rather than at full industrial maturity. AI-assisted HAZOP tools often demonstrate knowledge formalization or scenario generation, but not always long-term integration with industrial HAZOP workflows. Digital twin applications may demonstrate real-time monitoring or dynamic simulation, but not always explicit linkage to HAZOP knowledge. Hybrid models may demonstrate prediction or fault diagnosis, but not always safety-case validation. XAI methods may explain model outputs, but not always in process safety language suitable for operators and auditors.

For this reason, it is more accurate to describe the field as promising but not yet fully mature. The evidence supports the feasibility of individual components, but fewer studies demonstrate the full lifecycle chain: HAZOP knowledge formalization, live plant data integration, predictive modeling, explainable AI output, safeguard-status verification, human-in-the-loop decision support, and feedback into HAZOP revalidation.

7.3. Industrial Application Domains

Potential application domains for HAZOP-informed digital twins include reactors, distillation columns, heat exchangers, compressors, storage tanks, pipelines, utilities, and hydrogen or petrochemical safety systems [28,74]. These systems are relevant because they involve dynamic behavior, measurable process variables, known degradation mechanisms, and well-established HAZOP deviation patterns.

Reactors are particularly relevant because deviations in temperature, pressure, feed composition, cooling duty, or agitation can escalate rapidly. A HAZOP-informed digital twin could monitor thermal margins, reaction-rate indicators, cooling performance, and safeguard availability. Hybrid models may support early warning of thermal risk, while XAI can explain whether the developing scenario is linked to cooling loss, feed disturbance, or kinetic acceleration.

Distillation systems provide another important application area. Deviations such as high pressure, flooding, loss of separation, reboiler duty loss, condenser failure, or reflux malfunction are commonly considered in HAZOP studies. A digital twin can monitor pressure drop, tray or packing performance, temperatures, compositions, and control-loop behavior. Connecting these indicators to HAZOP knowledge could support earlier recognition of flooding, weeping, off-specification product, or pressure-related hazards.

Heat exchangers and utility systems are also suitable because degradation often develops gradually. Fouling, loss of cooling water, steam-side problems, or reduced heat-transfer performance may erode safety margins before alarms are triggered. A hybrid model can estimate heat-transfer degradation, while the HAZOP knowledge layer can relate loss of duty to downstream temperature, pressure, or reaction hazards.

Hydrogen systems and petrochemical units are increasingly relevant because they combine high-consequence hazards with growing interest in digital monitoring and safety analytics. Recent digital twin studies in hydrogen refueling and petrochemical safety show the potential for combining simulation, sensor data, and decision support [15,16]. However, these applications still need closer connection to formal process hazard analysis and lifecycle safety governance before they can be considered mature HAZOP-informed systems. Supplementary Table S7 extends this discussion with sector-specific implementation guidance for oil and gas offshore, petrochemical/refining, hydrogen energy, carbon capture, pharmaceuticals, and power generation, including the recommended pathway, priority actions, key barriers, and a five-year TRL target for each sector.

7.4. Implementation Barriers

Several barriers limit industrial deployment [28,29,56]. The first is data quality and availability. HAZOP-informed digital twins depend on reliable sensor data, consistent tag mapping, historian access, alarm records, maintenance information, and clear links between process variables and HAZOP nodes. Many plants have fragmented data systems, inconsistent naming conventions, missing contextual information, or limited access to safety-critical data.

The second barrier is knowledge formalization. HAZOP worksheets are often written for human readers, not machines. They may contain inconsistent terminology, incomplete causal descriptions, site-specific abbreviations, or vague recommendations. Transforming these records into a structured knowledge base requires expert review and careful standardization. Automated extraction can help, but it cannot fully replace domain validation.

The third barrier is model validation. Predictive models used for safety support should be validated under conditions that matter for process safety, including abnormal operation, start-up and shutdown, equipment degradation, sensor faults, and rare but credible hazardous scenarios. This is difficult because industrial data for severe abnormal events are limited, and deliberate testing of hazardous scenarios is often impossible.

The fourth barrier is organizational integration. A DT-AI system should fit into existing workflows for HAZOP, management of change, alarm management, maintenance, operating procedures, training, and incident learning. If the system is isolated from these workflows, it may become a digital dashboard rather than a safety-management tool.

The fifth barrier is human acceptance. Operators and engineers need to understand the system, trust it appropriately, and know how to act on its outputs. Poorly explained warnings, excessive alerts, or recommendations that conflict with operating experience can reduce trust. Conversely, overly confident explanations may encourage automation bias.

The sixth barrier is governance and cybersecurity. A digital twin connected to live plant data and safety-support logic should be protected against unauthorized access, data manipulation, model tampering, and uncontrolled updates. Model governance, audit trails, access control, and management-of-change procedures are therefore important for reliable deployment.

7.5. Maturity Comparison of the Four Pathways

Table 3 summarizes the relative maturity of the four pathways qualitatively. A detailed technology readiness level (TRL) assessment using the standard nine-point NASA/ESA scale [99] is provided in Supplementary Table S11 to preserve the evidence trail without overloading the main text. Both assessments are approximate: TRL assignments for software-intensive and knowledge-based systems carry inherent uncertainty and reflect the reviewed evidence rather than a formal certification exercise.

To make the qualitative maturity labels used in Table 3 and elsewhere in this review more transparent and less impressionistic, the following criteria are applied consistently: a pathway is labeled *strong* where the reviewed evidence includes quantitative validation in a relevant industrial or pilot environment; *moderate* where evidence is limited to academic case studies, laboratory-scale demonstration, or validation on a narrow process sub-domain; and *conceptual/developing* where a capability has been proposed or demonstrated only in simulation or benchmark settings with no published industrial or pilot validation. These same criteria are applied to the checkmark notation in Table 2 (✓ = covered; △ = partially covered; × = not covered) and were used to check the consistency of every classification in this review; entries that had previously implied full validation without a corresponding industrial-scale citation were revised downward accordingly.

Table 11 extends this maturity comparison with a practitioner-facing view: for each *technology pathway* (not specific commercial products), it summarizes the typical platform landscape, primary industry domain, implementation complexity, and enterprise scale at which the pathway is currently most plausible. This is an illustrative pathway-level comparison, not a systematic market or vendor review; named vendor and platform examples are illustrative rather than independently sourced or dated claims about specific products' capabilities, and readers seeking a rigorous competitive product analysis should consult primary vendor documentation directly.

7.6. Study-Level Evidence Maturity Assessment

The pathway-level assessment in Table 3 necessarily aggregates across many individual studies. To make the underlying evidence base transparent without extending the main text, Supplementary Table S1 presents a conservative study-level assessment of ten representative empirical or applied studies that can be matched directly to the bibliography. For each study it reports the pathway and application, data source, validation type, industrial scale, reported validation duration, approximate TRL, and a study-specific limitation. Review and framework papers are excluded from this table because they do not constitute study-level validation evidence. Where sustained validation duration is not reported, the table records "NR" rather than inferring duration from an industrial setting. The resulting evidence picture is cautious: several studies use genuine industrial documents or field settings, but none demonstrates long-term, multi-site validation of an integrated DT-AI-HAZOP system. This gap between prototype or bounded case study

evidence and sustained industrial validation is central to the conclusions of this review (see also Section 8).

Table 11. Illustrative comparative overview of DT–AI–HAZOP *technology pathways* by implementation context and organizational scale. This is not a systematic market or vendor review. The terms small, medium, and large are qualitative implementation contexts reflecting differences in process complexity, data infrastructure, specialist resources, and organizational capacity; they are not standardized industry classifications. Named vendor or platform examples are illustrative only and are not independently verified product-capability claims.

Pathway	Illustrative Platform Landscape	TRL Range	Implementation Complexity	Most Plausible Enterprise Scale
AI-assisted HAZOP (ontology, KG, NLP, LLM)	Ontology/knowledge graph tooling built on general graph-database platforms; LLM-assisted drafting via general-purpose LLM APIs with domain prompting.	2–7 (component-dependent)	Low–medium	Medium: Accessible with moderate investment once a HAZOP knowledge base exists; large facilities benefit most from cross-facility knowledge reuse.
Digital twin monitoring (five-layer architecture)	Commercial DT platforms integrated with plant DCS/historian systems.	2–6	High	Large: Most viable where mature DCS/historian infrastructure and dedicated OT/IT integration resources already exist.
Hybrid physics–data models (PINN, Bayesian, serial/parallel hybrids)	Research-stage frameworks and in-house implementations; limited commercial productization.	3–6	High	Large / research-active: Requires specialized modeling expertise currently concentrated in large or research-partnered organizations.
Explainable AI (SHAP, LIME, inherently interpretable models)	Open-source XAI libraries layered on existing ML models.	1–7 (method-dependent)	Low	Multiple scales: Potentially deployable as an overlay on existing models, making it a comparatively low-barrier entry point when a validated predictive model and suitable data pipeline already exist.

Note: TRL ranges reflect the spread across sub-components summarized in Supplementary Table S11; individual deployments will sit at a specific point within each range depending on scope. Vendor/platform examples are illustrative, not exhaustive, and do not constitute an endorsement.

7.7. Reported Quantitative Performance Metrics Across the Four Pathways

A persistent difficulty in evaluating DT–AI technologies for HAZOP enhancement is the scarcity of standardized, comparable performance metrics across the four pathways. Reported results include high accuracy in bounded AI-assisted HAZOP classification tasks [100], promising LLM worksheet-similarity scores but low valid-scenario ratios in open-ended generation [101], benchmark evidence for fault detection and hybrid modeling [95–97], selected PINN-based improvement metrics [102], and early XAI fault-diagnosis studies [83]. However, these metrics are not directly comparable across tasks, process types, datasets, or validation conditions. The detailed metric summary is therefore moved to Supplementary Table S10; the main finding is that live-plant false positive rates, false negative rates, detection latency, explanation fidelity, and operator-trust metrics remain insufficiently standardized for industrial HAZOP adoption [29,56,74,88].

7.8. Technology Readiness Level Assessment

Technology readiness levels (TRL 1–9) provide a standardized framework for assessing maturity from basic research to full operational deployment [99]. A detailed sub-component TRL assessment is provided in Supplementary Table S11. The main pattern is consistent: component-level technologies such as rule-based HAZOP support, individual ML models, XAI methods, and digital twin architectures are generally more mature than fully integrated HAZOP-specific systems. Rule-based HAZOP expert systems and general DT monitoring occupy the higher end of the reviewed maturity range, while LLM-based HAZOP generation, safeguard-status integration, and HAZOP-specific XAI remain earlier-stage. This gap between component maturity and integrated-system maturity defines the current research

landscape and motivates multidisciplinary work connecting process safety engineering, AI/ML, operational technology, and governance [29,56].

7.9. Critical Assessment of Industrial Readiness

The evidence reviewed in this section suggests that DT–AI-enhanced HAZOP is not a single ready-made technology but an emerging integration of several partially mature capabilities [28,29,56], each individually useful but most valuable when integrated in a disciplined safety-management framework. The main maturity gap is therefore systemic, not only technical: the field has many demonstrations of individual components but fewer examples of complete lifecycle systems connecting HAZOP worksheets, live data, predictive models, explanations, safeguard verification, and feedback into revalidation. This gap matters because process safety depends on the whole sociotechnical system’s performance, not an algorithm’s accuracy alone. A phased implementation approach, detailed in Section 8.7, is a realistic path to near-term adoption. While the four pathways show clear potential, their safe adoption requires careful attention to failure modes, validation evidence, organizational integration, cybersecurity, and regulatory confidence, discussed next.

8. Critical Discussion, Research Gaps, and Future Research Agenda

The preceding sections examined four complementary pathways for enhancing HAZOP through digital twin and artificial intelligence technologies: AI-assisted HAZOP, digital twin-based monitoring, hybrid physics–data models, and explainable AI. Taken together, these pathways suggest that future HAZOP enhancement is unlikely to depend on one technology alone. A major opportunity lies in integration: structured HAZOP knowledge provides the safety logic, digital twins provide the operational context, hybrid models provide predictive capability, and explainable AI provides the transparency needed for human trust and regulatory confidence.

However, the review also shows that DT–AI-enhanced HAZOP should not be interpreted as a replacement for conventional HAZOP. The core value of HAZOP remains expert-led, multidisciplinary, auditable reasoning about deviations from design intent. Digital technologies can strengthen this reasoning by improving knowledge retrieval, monitoring known deviations, forecasting hazardous trajectories, ranking plausible causes, and supporting decision making. They cannot remove the need for engineering judgment, plant-specific knowledge, safeguard evaluation, or organizational accountability.

This section synthesizes the main findings, cross-cutting failure modes, implementation requirements, and future research agenda.

8.1. Comparative Analysis: DT–AI-Enhanced vs. Conventional HAZOP

A direct comparison between DT–AI-enhanced and conventional HAZOP is useful for evaluating whether the proposed framework delivers practical value. Table 12 provides a structured head-to-head comparison across key performance and implementation dimensions. It is important to note that this comparison is qualitative and indicative: quantitative performance data from long-term industrial deployments remain limited, as most implementations are at the prototype or pilot scale (TRL 3–6). The comparison is therefore based on the evidence available in the reviewed literature.

Table 12. Comparative analysis of conventional HAZOP versus DT–AI-enhanced HAZOP across key performance and implementation dimensions. Ratings: ++ = strong advantage; + = advantage; ~ = comparable or context-dependent; - = disadvantage; – = significant disadvantage.

Dimension	Conventional HAZOP	DT–AI-Enhanced HAZOP	Notes
Hazard identification coverage	+	+	AI-assisted HAZOP can help retrieve historical scenarios and reduce omissions from inconsistent terminology; human expertise remains necessary for novel deviations [65,67,71].
Dynamic / real-time monitoring	–	++	Conventional HAZOP is periodic; DT-based monitoring can compare live plant state against HAZOP-identified deviation boundaries during operation [16,39,45].
Knowledge reuse and consistency	~	+	Ontologies and knowledge graphs enable systematic reuse of previous HAZOP knowledge; conventional practice relies on individual expert memory and unstructured documents [17–19].
Predictive capability	-	+	Hybrid physics–data models can support deviation forecasting and time-to-limit estimation before threshold exceedance; conventional HAZOP identifies credible scenarios prospectively but does not continuously predict their development during operation [51,79,80].
Transparency and auditability	++	~	Conventional HAZOP worksheets are inherently auditable; DT–AI systems need XAI and governance frameworks to approach equivalent auditability [25,85,86].
Regulatory acceptance	++	-	Conventional HAZOP is well-established in IEC 61882, OSHA PSM, and Seveso III; DT–AI augmentation lacks sector-specific regulatory guidance and requires proactive engagement [21–23].
Implementation cost and complexity	-	–	Both approaches require resources. Conventional HAZOP requires multidisciplinary team time, preparation, documentation, and follow-up, while DT–AI systems additionally require data infrastructure, sensors, model development, validation, cybersecurity, and governance [38,46].
Time to conduct HAZOP study	-	~	AI-assisted preparation, scenario retrieval, and worksheet generation can reduce study preparation time; however, expert review and validation time remains necessary [53,54,65].
Adaptability to plant changes	-	+	DT–AI systems can be updated as process conditions change; conventional HAZOP requires formal revalidation studies that are costly and time-consuming.
Human skill requirements	~	-	Conventional HAZOP requires strong process, operations, and facilitation expertise. DT–AI systems add new competency requirements in data science, AI governance, model validation, cybersecurity, and digital safety assurance [24,47].
Risk of automation complacency	~	-	Conventional HAZOP has human-factor risks such as groupthink, incomplete participation, and documentation bias. DT–AI systems add automation complacency and overtrust risks, which call for governance, training, uncertainty communication, and clear advisory framing [24,47].

Note: Ratings are qualitative author assessments of relative capability and implementation burden based on the reviewed evidence. They do not represent measured effect sizes or results from direct controlled comparisons. DT–AI-enhanced HAZOP is interpreted as an augmentation of, rather than a replacement for, conventional expert-led HAZOP.

8.2. Industrial Case Study Synthesis

Although integrated DT–AI HAZOP systems at the industrial scale remain rare, the reviewed literature contains evidence of partial implementations that illustrate the practical potential and boundary conditions of individual pathway components. Four representative case studies are synthesized below, each structured consistently as problem, implemented solution, results and metrics, and limitations identified, so that positive outcomes are reported alongside the practical constraints encountered.

Case Study 1: Hydrogen Refueling Station Digital Twin Prototype (An et al. [46]).

Problem: Early leak and abnormal-condition detection for hydrogen refueling, where high-consequence releases can occur rapidly. **Solution:** A digital twin prototype combining previously collected station data, a CNN-based risk-prediction model, and 3D visualization;

real-time IoT collection and automated control were outside the reported study. *Results:* The CNN achieved an F1 score of 0.993 and 68 ms prediction time on the study dataset, supporting one-second decision updates in the prototype. *Limitations:* The abnormal-condition samples were artificially generated, no live IoT trial was performed, cross-station generalizability was not validated, and OT cybersecurity was not evaluated.

Case Study 2: AIMNET Gas Emission Monitoring Digital Twin (Zhou et al. [103]).

Problem: Methane monitoring across a distributed industrial site with sparse fixed sensors. *Solution:* An IoT-empowered digital twin combining a low-power sensor network with a physics-based weather–gas transport model around a wastewater treatment plant. *Results:* The preliminary field demonstration captured methane-event dynamics and expanded spatial interpretation through tiered simulation; a standardized incident-reduction or response-time metric was not reported. *Limitations:* Preprint evidence, dependence on meteorological input quality, sensor-network maintenance burden, and no published integration with formal HAZOP records.

Case Study 3: HAZOP-Integrated Dynamic Simulation for an Exothermic Reaction (Berdouzi et al. [104]). *Problem:* Static HAZOP review cannot capture dynamic consequence escalation in an exothermic semi-batch reaction, where severity depends on timing and duration. *Solution:* Aspen Plus Dynamics simulation integrated with HAZOP guide-word review and risk matrices. *Results:* The method provided time-dependent consequence assessment for runaway-reaction scenarios; no operational accident-reduction or response-time metric was reported. *Limitations:* Design-stage focus, no real-time operational monitoring, and need for dynamic-simulation expertise not universally available within HAZOP teams.

Case Study 4: Industrial Safety Knowledge Graph for an Indirect Coal Liquefaction Process (Wang et al. [105]). *Problem:* Dispersed HAZOP and process safety knowledge across many process units. *Solution:* An industrial safety knowledge graph built with HAZOP as the organizing carrier, using deep learning-based entity recognition. *Results:* The case demonstrated structured knowledge retrieval and improved entity recognition on the studied corpus; no operational safety-outcome metric was reported. *Limitations:* Process specificity, substantial expert effort for extension, and residual safety-knowledge gaps from missed or misclassified entities.

These case studies collectively illustrate that the pathway from component demonstration to integrated HAZOP-informed digital twin systems requires not only technical integration, but also systematic validation, governance, and organizational embedding. A key lesson across the cases is that digital tools add value when they are explicitly aligned with HAZOP structure, expert judgment, and the specific decision needs of process safety teams. Even the strongest reported model result (the F1 score of 0.993 in Case Study 1) comes from a single-dataset prototype using artificially generated abnormal samples and has not been independently replicated across sites.

8.3. Why Industrial Pilots Stall: A Barrier Analysis

The case studies above, and the broader evidence base summarized in Supplementary Table S1, share a common pattern: technically successful pilots that have not progressed to sustained, multi-site industrial deployment. Drawing on evidence across the reviewed literature, five cross-cutting barrier categories explain why this progression stalls.

1. **Data infrastructure gaps.** Most facilities lack the sensor density, historian configuration, and data quality that DT–AI systems assume. The case studies above illustrate this directly: AIMNET’s performance depends on meteorological input quality and a maintained distributed sensor network (Case Study 2), and none of the four cases report integration with existing DCS/historian infrastructure as a solved problem.

2. **IT/OT integration complexity.** Connecting DT platforms to DCS, SIS, and SCADA environments involves substantial systems-integration cost and cybersecurity risk; none of the four case studies above address OT-connected cybersecurity in detail in the published work, indicating that this aspect remains early-stage.
3. **Organizational inertia and ROI uncertainty.** Process safety investment competes with production-optimization investment, and the commercial case for AI-enhanced HAZOP is difficult to quantify under current regulatory frameworks, which do not yet formally credit DT–AI monitoring in safety-case arguments.
4. **Validation and liability uncertainty.** Without regulatory guidance on AI-advisory systems (Section 8.14), organizations are reluctant to accept liability for AI-influenced safety decisions, which discourages the sustained, multi-site validation needed to move a pilot toward routine operational use.
5. **Skills and capability gaps.** Combining process safety domain expertise with AI/data-science capability is rare and expensive; Case Study 3 illustrates this directly, since HAZOP-integrated dynamic simulation requires simulation expertise not universally available within HAZOP teams.

These barriers are not primarily technical; they help explain why the field’s technology-readiness levels remain concentrated at TRL 3–6 despite two decades of methodological progress (Supplementary Table S11), and they motivate the phased implementation roadmap in Section 8.

8.4. Synthesis of the Four Pathways

The four pathways address different limitations of conventional HAZOP. AI-assisted HAZOP improves knowledge formalization, retrieval, and consistency, but remains most defensible as expert support rather than autonomous hazard analysis [65,67,71]. Digital twin architectures connect HAZOP knowledge to changing plant conditions, helping identify developing deviations, shrinking margins, and safeguard assumptions that need checking [39,45,74]. Hybrid physics–data models add predictive safety analytics, including trajectory forecasting, time-to-limit estimation, and cause discrimination, while introducing validation, uncertainty, drift, and rare-event generalization issues [53–55]. XAI addresses the human and regulatory side by translating model outputs into HAZOP language: deviation, cause, consequence, safeguard, and recommendation [85–87]. The pathways are complementary; the main task is to integrate them into a coherent safety-management framework.

8.5. Cross-Cutting Failure Modes

DT–AI-enhanced HAZOP introduces new opportunities, but also new failure modes [48,56,88]: limited knowledge-base coverage, model drift, data-quality degradation, misleading explanation, automation bias, and organizational detachment. If the HAZOP knowledge base omits a credible scenario, the digital twin may not monitor it and the AI system may not retrieve it [48,56]. Stale models can create false confidence, poor data governance can weaken safety-support quality, and feature-attribution explanations can appear clear without explaining causal pathways or safeguard implications. Operators may over-rely on confident outputs, especially outside the validated domain, and technically impressive tools can become isolated dashboards if disconnected from HAZOP revalidation, MOC, alarm management, maintenance, procedures, and training. Supplementary Table S4 pairs each of these failure modes with its possible safety consequence, a corresponding detection method, and a mitigation strategy.

8.6. Implementation Principles for Safe Adoption

The review supports seven practical principles for safe adoption [29,48,56]: augmentation rather than replacement; alignment with HAZOP structure; validation before reliance; uncertainty communication; human-in-the-loop governance; lifecycle integration; and cybersecurity/configuration control. In practice, DT-AI systems should support expert-led HAZOP, express outputs in HAZOP language, validate models and knowledge bases against expert-reviewed abnormal scenarios, communicate confidence and operating-domain limits, preserve professional accountability, feed results back into MOC and revalidation, and protect live-data/model infrastructure against unauthorized changes.

8.7. Phased Implementation Roadmap

The transition from conventional HAZOP to DT-AI-enhanced lifecycle hazard intelligence is a multi-year organizational transformation, not a single technology deployment. Table 13 presents a four-phase roadmap addressing data quality, sensor reliability, DCS/historian integration, alarm management, management of change (MOC), functional safety interfaces, cybersecurity, operator training, validation responsibility, and regulatory acceptance. The roadmap is a general framework; specific timelines and priorities will vary by facility type, sector, regulatory context, and organizational capability, and it should be read alongside the barrier analysis in Section 8.3, which explains why progression between phases is often slower than technology readiness alone would suggest. Supplementary Table S5 reproduces this roadmap with an added column mapping each phase explicitly to the key implementation dimensions it addresses (data quality, sensor reliability, alarm management, functional safety interfaces, cybersecurity, MOC, skills, and regulatory acceptance).

Table 13. Phased DT-AI-HAZOP implementation roadmap.

Phase	Duration	Key Activities	Success Criteria
Phase 1: Readiness assessment	Months 1–6	HAZOP document audit; data-infrastructure gap analysis; DCS/historian coverage mapping; cybersecurity baseline (IEC 62443); regulatory pre-engagement.	Documented gap analysis; go/no-go investment decision.
Phase 2: Supervised pilot	Months 6–24	Single-unit DT/AI deployment; alarm-management baseline (EEMUA 191); advisory-mode only, with all outputs reviewed by qualified engineers; MOC integration for model updates; operator training program.	Detection rate meets pre-agreed threshold; false-alarm performance meets site-defined targets aligned with the alarm-management principles and performance metrics of EEMUA 191; operator acceptance survey.
Phase 3: Controlled scale-up	Years 2–4	Multi-unit rollout; SIS/IEC 61511 boundary definition; functional safety interface documentation; cybersecurity program aligned with applicable IEC 62443 requirements, with a target security level—such as SL 2 where justified—established through site-specific risk assessment; competency-maintenance requirements.	Predefined safety-performance and validation criteria are met, with no unresolved safety-critical findings during supervised evaluation; incident-free scale-up; regulatory notification where required.
Phase 4: Enterprise governance	Year 4+	Validation monitoring; periodic independent audit; AI accountability framework; regulator engagement on AI HAZOP guidance; technology refresh cycle.	Annual independent audit; documented regulatory engagement and evidence of compliance or acceptance where required by the applicable jurisdiction; sustained operator trust metrics.

Note: Responsible parties—phases 1–2: process safety engineer (lead), IT/OT integration, instrumentation; Phases 3–4: process safety manager (lead), legal/compliance, regulator liaison; validation responsibility remains with qualified professionals throughout. Durations and success criteria are indicative planning ranges, not prescriptive requirements, and should be set through site-specific risk assessment and applicable standards.

8.8. Risk and Barrier Assessment for DT–AI–HAZOP Adoption

Beyond the technology-focused barriers in Section 8.3, Supplementary Table S6 presents thirteen risks across technical, organizational, economic, security, and scalability categories, each paired with a mitigation measure. The priority risks shaping the recommendations are model drift, rare-event data scarcity, deskilling of engineering judgment, high upfront infrastructure cost, OT network exposure, and limited transferability from large continuous processes to batch operations or smaller facilities. DT–AI approaches are most likely to work effectively in large continuous processes with mature data infrastructure and process safety engineering depth; site-specific risk assessments remain necessary for actual deployment.

8.9. Research Gaps

Several research gaps remain [29,48,56]. The first is the lack of validated end-to-end HAZOP-informed digital twin frameworks. Many studies address individual components, such as ontology-based HAZOP, digital twin monitoring, fault diagnosis, or XAI [29]. Fewer studies demonstrate the complete pathway from HAZOP knowledge formalization to live data integration, predictive modeling, explanation, safeguard verification, and feedback into revalidation.

The second gap is limited industrial validation. Simulation studies and proof-of-concept demonstrations are valuable, but process safety adoption requires evidence under real operating conditions. Future work should evaluate DT–AI systems across different process units, operating modes, abnormal scenarios, and organizational contexts.

The third gap is the treatment of safeguards. Many digital safety studies focus on detecting abnormal conditions, but fewer explicitly evaluate whether safeguards are available, independent, effective, bypassed, degraded, or already demanded. Since safeguards are central to HAZOP and protection-layer thinking, this gap is important.

The fourth gap is explainability in HAZOP language. General XAI methods are not enough. Future research should develop explanation methods that directly map model outputs to deviation, cause, consequence, safeguard, and recommended verification.

The fifth gap is uncertainty and model-boundary communication. DT–AI systems should inform users when the model is extrapolating, when data quality is poor, or when predictions are uncertain. This remains underdeveloped in many safety-support prototypes.

The sixth gap is governance for LLM use in HAZOP. LLMs may support drafting, retrieval, and summarization, but their use in safety-critical workflows requires strict controls for hallucination, source traceability, review, versioning, and accountability.

The seventh gap is human-factor validation. Future work should evaluate how operators and engineers actually interact with DT–AI explanations during abnormal situations. Usability, workload, trust calibration, alarm fatigue, and automation bias should be studied alongside model accuracy.

8.10. Future Research Agenda

A staged research agenda is needed to move the field from fragmented demonstrations toward reliable industrial deployment [28,29]. This agenda should address human factors, ethics, and governance requirements [56,88]. Figure 6 summarizes this agenda by organizing future work into short-term, medium-term, and long-term horizons, supported by cross-cutting requirements for validation, human factors, governance, and cybersecurity.

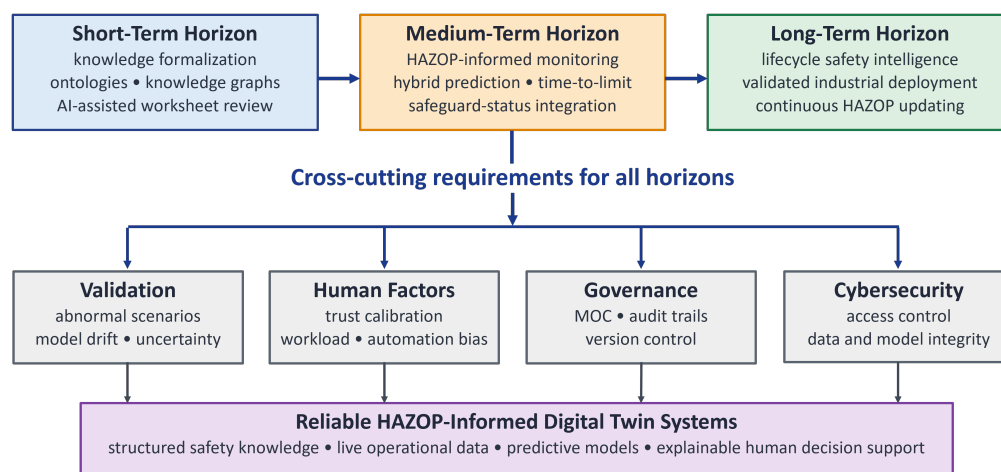


Figure 6. Future research agenda for DT-AI-enhanced HAZOP. The agenda progresses from short-term knowledge formalization, through medium-term HAZOP-informed monitoring and predictive analytics, toward long-term lifecycle safety intelligence. Validation, human factors, governance, and cybersecurity are shown as cross-cutting requirements for all horizons.

In the short term, research should focus on knowledge formalization and expert-support functions. This includes standardizing HAZOP terminology, building reusable ontologies, developing knowledge graphs from HAZOP and incident data, and evaluating AI-assisted worksheet review under expert supervision.

In the medium term, research should focus on HAZOP-informed monitoring and predictive analytics. This includes mapping HAZOP deviations to measurable indicators, developing hybrid models for selected high-risk units, validating time-to-limit prediction, and integrating model outputs with safeguard-status information. Demonstrations should move beyond idealized case studies toward realistic plant data, degraded conditions, and abnormal operating scenarios.

In the longer term, research should focus on lifecycle safety intelligence. This means developing systems that connect HAZOP knowledge, digital twin monitoring, predictive models, explainable decision support, incident learning, management of change, and revalidation over the plant lifecycle. The goal should not be autonomous HAZOP, but a disciplined human-centered safety-support ecosystem that keeps hazard knowledge alive throughout the plant lifecycle.

Future research should also develop evaluation metrics that go beyond prediction accuracy. Important metrics include omission rate for credible hazards, false advisory rate, warning lead time, explanation usefulness, operator response quality, safeguard-verification effectiveness, model drift detection, and contribution to HAZOP revalidation. These metrics would better reflect the real needs of process safety.

8.11. Critical Position of This Review

The overall position of this review is deliberately balanced. DT-AI technologies offer meaningful opportunities to enhance HAZOP, but their value depends on disciplined integration with process safety practice. The main contribution is not full automation; it is the ability to make HAZOP knowledge more structured, dynamic, predictive, explainable, and reusable [71,96,106].

This position differs from overly optimistic views that present AI as a replacement for expert hazard analysis. It also differs from overly conservative views that treat digital methods as irrelevant to process safety. The more realistic position is that AI and digital twins can become valuable safety-support layers when they are transparent, validated, governed, and aligned with expert-led HAZOP practice [88,106,107].

The future of HAZOP enhancement therefore depends less on whether AI can “perform HAZOP” and more on whether AI and digital twins can help process safety teams maintain a living understanding of hazards as plants change over time. This lifecycle perspective is the central contribution of the DT–AI-enhanced HAZOP framework proposed in this review [71,97].

8.12. Limitations of This Review

This review has several limitations that should be considered when interpreting its findings. *Scope and methodology*: This is a thematic critical literature review, not a systematic review; source selection reflects the author’s judgment rather than a pre-registered protocol, and readers seeking pathway-specific systematic coverage should consult dedicated systematic reviews. *Uneven evidence base*: AI-assisted HAZOP has a substantially longer research history than hybrid physics–data modeling or XAI for HAZOP decisions specifically, so cross-pathway maturity comparisons (Table 3) should be interpreted with this asymmetry in mind. *Simulation and prototype bias*: Much of the digital twin monitoring and hybrid modeling evidence is simulation-based or prototype-scale (Supplementary Table S1); real industrial performance may differ substantially, and long-term deployment evidence is almost entirely absent, as discussed throughout Section 8. *Rapid technological change*: AI and digital twin capabilities evolve quickly; LLM capability and XAI maturity assessments in particular may date faster than other parts of this review. *Sector and regional variability*: Evidence is concentrated in oil-and-gas and petrochemical applications; findings may not generalize to all sectors or regulatory environments. *Single-author perspective*: Single-author synthesis introduces selection, emphasis, and interpretive biases not mitigated by independent screening; consistent critical standards and explicit uncertainty acknowledgment throughout this review are the primary countermeasures applied.

8.13. Ethical, Workforce, and Regulatory-Engagement Considerations

The deployment of AI and digital twin technologies in process safety raises ethical and societal questions beyond technical performance, especially accountability, transparency, workforce impact, and social license [56,84,108]. For HAZOP applications, the most conservative position is human-in-the-loop accountability with strong auditability: AI may generate or retrieve candidate scenarios, but competent professionals should verify outputs and retain responsibility for final safety decisions [56,88,92,93,108]. Meaningful transparency also requires documented data provenance, a validated operating envelope, traceable outputs, and performance reporting that engineers can interpret in safety terms, not merely mathematical terms [83,84,108,109]. Responsible deployment should include upskilling, participatory design with operators and process safety engineers, and early regulatory engagement, because dedicated AI-assisted-HAZOP guidance remains limited [56,88,92].

The ethical position adopted throughout this review is that DT–AI technologies should strengthen, not dilute, process safety responsibility: AI may assist with retrieval, generation, monitoring, prediction, and explanation, but final safety judgment should remain with competent human professionals within established safety-management systems, evaluated not only on technical performance but on whether it preserves professional accountability, safety culture, and public confidence in high-hazard operations.

8.14. Verification, Validation, and Functional Safety Standards

The deployment of AI and digital twin components in HAZOP-related safety functions raises V&V questions not fully addressed by traditional process safety standards. IEC 61508, the foundational functional-safety standard defining Safety Integrity Levels (SIL 1–4) [110], presents four specific issues for AI/ML components [24,111]: its V&V assumes deterministic, fully specifiable software behavior, whereas ML models are probabilistic

and cannot be fully characterized by finite test cases; its hardware-reliability framework does not accommodate ML-specific failure classes (distributional shift, adversarial inputs, hallucination); it does not address training-data quality, coverage, or bias as safety artifacts; and its V-model lifecycle assumes stable requirements, whereas ML systems require monitoring, retraining, and revalidation [24,92,109,111]. Ongoing IEC 61508 revisions are beginning to address AI/ML but remain incomplete; practitioners are advised to apply the standard as a baseline supplemented with AI-specific V&V [24,92,109,111]. UL 4600, developed for autonomous products, offers a useful safety-case analogy: a structured argument, with supporting evidence, that the system is acceptably safe for its intended use, requiring explicit ML failure-mode analysis, a defined operational design domain (process types, operating modes, data-quality conditions, decision-support boundaries), and ongoing monitoring for degradation or operation outside the validated domain [112], informed by the broader AI/ML safety-assurance literature [92,109,111,113].

Drawing on both standards and the emerging AI safety literature, a layered V&V framework is proposed with four complementary layers [24,92,109,111–113]: (1) *formal and analytical verification* of deterministic components (rule engines, ontology reasoners), with bounded input–output verification for ML components; (2) *statistical testing and benchmarking* on held-out datasets representative of rare and abnormal scenarios, with per-condition performance metrics and adversarial testing [24,100,109]; (3) *process-specific validation* of AI-generated HAZOP outputs against expert-led studies on the same systems, assessing both false negatives and false alarms [29,56,92]; and (4) *operational monitoring and lifecycle management*, with drift/degradation monitoring, defined retraining and revalidation triggers, and auditability of model versions, training data, and deployment decisions [78, 84,92,111,113]. Figure 7 summarizes this layered V&V framework, showing how formal verification, statistical testing, process-specific validation, and operational monitoring combine to support IEC 61508-aligned assurance and UL 4600-inspired safety-case evidence across the lifecycle.

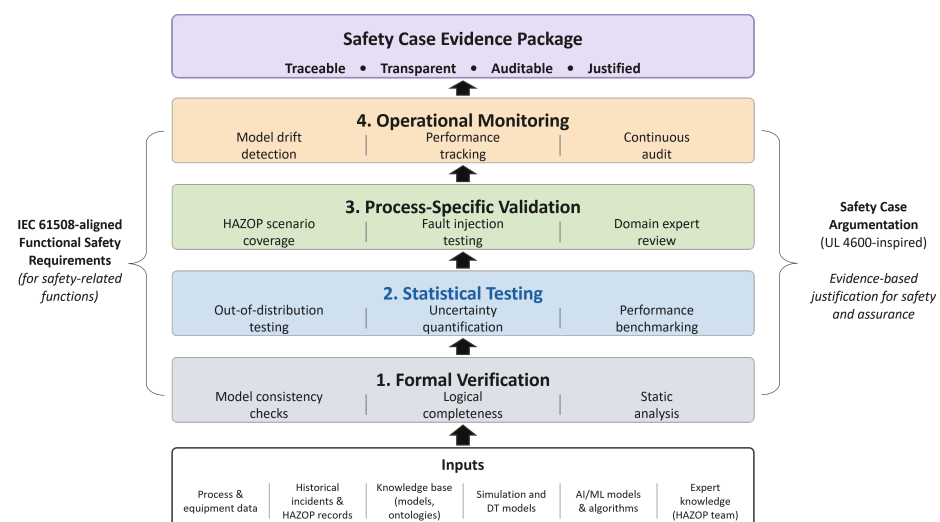


Figure 7. Layered V&V framework for DT-AI HAZOP systems. Formal verification, statistical testing, process-specific validation, and operational monitoring combine to support IEC 61508-aligned assurance and UL 4600-inspired safety-case evidence across the lifecycle.

A particularly demanding requirement is *coverage*: reducing the likelihood that an AI system omits safety-significant deviations. Traditional HAZOP manages this concern through systematic guide-word application to every node, whereas AI-generated deviation lists do not provide the same level of assurance without additional mechanisms [29,48,56]. Proposed approaches include independent expert verification against the full guide-word

checklist, ensemble methods flagging any deviation identified by at least one method, coverage-tracking knowledge bases, and uncertainty quantification triggering enhanced review for low-confidence nodes [48,56,109,111]. Each helps but none fully solves the problem; coverage assurance remains an open research problem for AI-assisted HAZOP [29,48,92].

9. Conclusions and Recommendations

This review critically examined how digital twin and artificial intelligence technologies can support the enhancement of conventional HAZOP in process safety [29,56]. Conventional HAZOP remains one of the most important methods for systematic hazard identification because it provides a structured, expert-led, and auditable way to examine deviations from design intent. However, its traditional implementation is often periodic, document-centered, and strongly dependent on expert experience. These characteristics limit its ability to reflect operational drift, equipment degradation, process modifications, changing feedstocks, and emerging abnormal conditions between formal review cycles.

The central conclusion of this review is that DT–AI technologies should not be viewed as replacements for HAZOP [48,56]. Rather, they should be designed as decision-support layers that strengthen expert-led process safety practice. The most credible direction is not autonomous HAZOP, but HAZOP-informed lifecycle safety intelligence. In this view, HAZOP knowledge is formalized, connected to operational data, supported by predictive models, explained to human users, and updated through learning from plant operation.

The review identified four complementary technology pathways: AI-assisted HAZOP for knowledge capture and scenario retrieval; digital twin architectures for linking HAZOP knowledge to live plant data and operator support; hybrid physics–data models for deviation forecasting, time-to-limit estimation, cause ranking, and uncertainty communication [47,76,78]; and explainable AI for making outputs understandable, reviewable, auditable, and acceptable to operators, engineers, safety personnel, and regulators [48,49,56,62]. Together, these pathways can help move HAZOP from a static worksheet toward an operational safety-support resource, but only under expert validation and governance.

The review also showed that the field is promising but not yet fully mature [29]. Many studies demonstrate individual components, such as ontology-based HAZOP, process safety knowledge graphs, digital twin monitoring, fault diagnosis, physics-informed modeling, or XAI. Fewer studies demonstrate integrated lifecycle systems that connect HAZOP knowledge formalization, live data acquisition, predictive modeling, safeguard-status monitoring, explainable decision support, and feedback into HAZOP revalidation [28,29,56]. This integration gap remains a major barrier to industrial adoption.

Several practical recommendations follow. Organizations should adopt DT–AI tools gradually, beginning with low-risk expert-support functions, such as knowledge retrieval, terminology standardization, consistency checking, and document review [29,56]. Digital twin monitoring should be explicitly linked to HAZOP deviation–cause–consequence–safeguard logic rather than used only as generic visualization or anomaly detection [28]. Predictive models should include validation boundaries, uncertainty communication, and model-drift monitoring before use in safety-related decision support [78,82]. AI-generated outputs, especially LLM outputs, should remain advisory unless formally validated and governed, with final hazard identification, safeguard judgment, and risk acceptance retained by competent professionals. Explainability should be expressed in process safety language: deviations, causes, consequences, safeguards, uncertainty, and recommended verification [56,91].

Future research should focus on better-validated end-to-end HAZOP-informed digital twin systems [28,29,56]. Priority areas include standardized HAZOP ontologies, process safety knowledge graphs, hybrid models for high-risk unit operations, uncertainty-aware early warning, safeguard-status monitoring, HAZOP-specific explainability, human factors validation, cybersecurity, and governance frameworks for AI-supported safety decisions. Research should also develop evaluation metrics that reflect process safety needs, including warning lead time, omission rate for credible hazards, explanation usefulness, safeguard-verification effectiveness, and contribution to HAZOP revalidation.

Overall, DT–AI-enhanced HAZOP should be understood as an evolution of process safety practice rather than a replacement of established methods [48,56]. Its value lies in keeping HAZOP knowledge alive throughout the plant lifecycle. When implemented with validation, explainability, governance, and human oversight, digital twins and artificial intelligence can help process safety teams move from periodic hazard review toward more lifecycle-oriented safety management [28,29]. Human factors and trust considerations remain important for successful adoption [56,88].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr14142334/s1>. Table S1: Study-Level Evidence Maturity Assessment; Table S2: Hybrid Physics–Data Modeling Approaches for Predictive Process Safety; Table S3: XAI Methods for Process Safety Applications; Table S4: Failure Modes, Safety Consequences, Detection Methods, and Mitigation Strategies for DT–AI–HAZOP Systems; Table S5: Phased DT–AI–HAZOP Implementation Roadmap; Table S6: DT–AI–HAZOP Implementation Risk Register; Table S7: Sector-Specific DT–AI–HAZOP Implementation Guidance; Table S8: Comparison of AI-Assisted HAZOP Approaches; Table S9: Literature Search Verification Log; Table S10: Reported Quantitative Performance Metrics; Table S11: Detailed TRL Assessment. No separate reference list is provided in the supplementary file; sources named in the supplementary tables are already cited in the main manuscript and included in the main reference list.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The author thanks the Chemical Engineering Department, Jubail Industrial College, for institutional support.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ANN	Artificial neural network
BERT	Bidirectional Encoder Representations from Transformers
BN	Bayesian network
BP	British Petroleum
CCPS	Center for Chemical Process Safety
CFD	Computational fluid dynamics
COMAH	Control of Major Accident Hazards
CSB	Chemical Safety Board (US)
DCS	Distributed control system
DT	Digital twin
ESA	European Space Agency
ESD	Emergency shutdown system
EU	European Union

F1	F1 score
FTA	Fault tree analysis
GPT	Generative Pre-trained Transformer
HAZOP	Hazard and operability study
HMI	Human–machine interface
IEC	International Electrotechnical Commission
IoT	Internet of Things
ISO	International Organization for Standardization
KG	Knowledge graph
LIME	Local Interpretable Model-agnostic Explanations
LLM	Large language model
ML	Machine learning
ML-FMEA	Machine Learning Failure Mode and Effects Analysis
MOC	Management of change
NASA	National Aeronautics and Space Administration
NLP	Natural language processing
ODD	Operational design domain
OWL	Web Ontology Language
P&ID	Piping and instrumentation diagram
PEM	Proton-exchange membrane
PHA	Process hazard analysis
PINN	Physics-informed neural network
PSM	Process safety management
RAG	Retrieval-augmented generation
RDF	Resource description framework
SHAP	Shapley Additive Explanations
SIL	Safety integrity level
SPARQL	SPARQL Protocol and RDF Query Language
TEP	Tennessee Eastman process
TRL	Technology readiness level
V&V	Verification and validation
XAI	Explainable artificial intelligence
XFDDC	Explainable Fault Detection and Diagnosis Classifier

References

1. U.S. Chemical Safety and Hazard Investigation Board. *Investigation Report: Refinery Explosion and Fire, BP Texas City, Texas, March 23, 2005*; Technical Report Report No. 2005-04-I-TX; U.S. Chemical Safety and Hazard Investigation Board: Washington, DC, USA, 2007.
2. Buncefield Major Incident Investigation Board. *The Buncefield Incident 11 December 2005: The Final Report of the Major Incident Investigation Board*; Technical report; Health and Safety Executive: London, UK, 2008.
3. Health and Safety Executive. *Buncefield: Why Did It Happen? The Underlying Causes of the Explosion and Fire at the Buncefield Oil Storage Depot, Hemel Hempstead, Hertfordshire on 11 December 2005*; Technical report; Health and Safety Executive: Sudbury, UK, 2011.
4. Hopkins, A. *Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout*; CCH Australia: Sydney, Australia, 2012.
5. National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling. *Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling*; Technical report; U.S. Government Printing Office: Washington, DC, USA, 2011.
6. U.S. Chemical Safety and Hazard Investigation Board. *Final Investigation Report: Chevron Richmond Refinery Pipe Rupture and Fire*. Technical Report Report No. 2012-03-I-CA; U.S. Chemical Safety and Hazard Investigation Board: Washington, DC, USA, 2015.
7. Yoshida, S. Massive Fire Incident of Aboveground Oil Storage Tanks in Jaipur, India. *J. High Press. Inst. Jpn.* **2013**, *51*, 53–59. <https://doi.org/10.11181/hpi.51.53>.
8. Center for Chemical Process Safety. *Guidelines for Hazard Evaluation Procedures*, 3rd ed.; AIChE/Wiley: Hoboken, NJ, USA, 2008.
9. IEC 61882:2016; Hazard and Operability Studies (HAZOP Studies)—Application Guide. International Electrotechnical Commission: Geneva, Switzerland, 2016.

10. Dunj6, J.; Fthenakis, V.; V6lchez, J.A.; Arnaldos, J. Hazard and Operability (HAZOP) Analysis: A Literature Review. *J. Hazard. Mater.* **2010**, *173*, 19–32. <https://doi.org/10.1016/j.jhazmat.2009.08.076>.
11. Baybutt, P. A Critique of the Hazard and Operability (HAZOP) Study. *J. Loss Prev. Process Ind.* **2015**, *33*, 52–58. <https://doi.org/10.1016/j.jlp.2014.11.010>.
12. Center for Chemical Process Safety. *Guidelines for Revalidating a Process Hazard Analysis*; AIChE/Wiley: New York, NY, USA, 2022.
13. Venkatasubramanian, V. The Promise of Artificial Intelligence in Chemical Engineering: Is It Here, Finally? *AIChE J.* **2019**, *65*, 466–478. <https://doi.org/10.1002/aic.16489>.
14. Lee, J.; Davari, H.; Singh, J.; Pandhare, V. Industrial Artificial Intelligence for Industry 4.0-Based Manufacturing Systems. *Manuf. Lett.* **2018**, *18*, 20–23. <https://doi.org/10.1016/j.mfglet.2018.09.002>.
15. Grieves, M.; Vickers, J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems*; Kahlen, F.J., Flumerfelt, S., Alves, A., Eds.; Springer: Cham, Switzerland, 2016; pp. 85–113. https://doi.org/10.1007/978-3-319-38756-7_4.
16. Bevilacqua, M.; Bottani, E.; Ciarapica, F.E.; Costantino, F.; Di Donato, L.; Ferraro, A.; Mazzuto, G.; Monteriù, A.; Nardini, G.; Ortenzi, M.; et al. Digital Twin Reference Model Development to Prevent Operators' Risk in Process Plants. *Sustainability* **2020**, *12*, 1088. <https://doi.org/10.3390/su12031088>.
17. Single, J.I.; Schmidt, J.; Denecke, J. Ontology-Based Computer Aid for the Automation of HAZOP Studies. *J. Loss Prev. Process Ind.* **2020**, *68*, 104321. <https://doi.org/10.1016/j.jlp.2020.104321>.
18. Single, J.I.; Schmidt, J.; Denecke, J. Computer-Aided HAZOP: Ontologies and AI for Hazard Identification and Propagation. In *Proceedings of the Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 48, pp. 1783–1788. <https://doi.org/10.1016/B978-0-12-823377-1.50298-6>.
19. Mao, S.; Zhao, Y.; Chen, J.; Wang, B.; Tang, Y. Development of Process Safety Knowledge Graph: A Case Study on Delayed Coking Process. *Comput. Chem. Eng.* **2020**, *143*, 107094. <https://doi.org/10.1016/j.compchemeng.2020.107094>.
20. Kletz, T.A. *HAZOP and HAZAN: Identifying and Assessing Process Industry Hazards*, 4th ed.; Institution of Chemical Engineers: Rugby, UK, 1999.
21. Occupational Safety and Health Administration. *Process Safety Management of Highly Hazardous Chemicals*, 29 CFR 1910.119; Code of Federal Regulations; Office of the Federal Register: Washington, DC, USA, 1992.
22. European Parliament and Council of the European Union. Directive 2012/18/EU on the Control of Major-Accident Hazards Involving Dangerous Substances. *Off. J. Eur. Union* **2012**, *197*, 1–37.
23. Health and Safety Executive. *The Control of Major Accident Hazards Regulations 2015*; UK Statutory Instruments 2015 No. 483; The Stationery Office (TSO): London, UK, 2015.
24. Schwalbe, G.; Schels, M. A Survey on Methods for the Safety Assurance of Machine Learning Based Systems. In *Proceedings of the 10th European Congress on Embedded Real Time Software and Systems (ERTS 2020)*, Toulouse, France, 29–31 January 2020.
25. Barredo Arrieta, A.; D6az-Rodr6guez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>.
26. Mohseni, S.; Zarei, N.; Ragan, E.D. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Trans. Interact. Intell. Syst.* **2021**, *11*, 1–45. <https://doi.org/10.1145/3387166>.
27. Khoo, T.L.; Lee, T.S.; Bee, S.T.; Ma, C.; Zhang, Y.Y. A Comparative Review of Large Language Models in Engineering with Emphasis on Chemical Engineering Applications. *Processes* **2025**, *13*, 2680. <https://doi.org/10.3390/pr13092680>.
28. Mane, S.; Dhote, R.R.; Sinha, A.; Thirumalaiswamy, R. Digital Twin in the Chemical Industry: A Review. *Digit. Twins Appl.* **2024**, *1*, 118–130. <https://doi.org/10.1049/dgt2.12019>.
29. Elhosary, E.; Moselhi, O. Automation for HAZOP Study: A State-of-the-Art Review and Future Research Directions. *J. Inf. Technol. Constr.* **2024**, *29*, 750–777. <https://doi.org/10.36680/j.itcon.2024.033>.
30. Mannan, S. *Lees' Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control*, 4th ed.; Butterworth-Heinemann: Oxford, UK, 2012.
31. Crawley, F.; Tyler, B. *HAZOP: Guide to Best Practice*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2015.
32. Vaidhyanathan, R.; Venkatasubramanian, V. Digraph-Based Models for Automated HAZOP Analysis. *Reliab. Eng. Syst. Saf.* **1995**, *50*, 33–49. [https://doi.org/10.1016/0951-8320\(95\)00052-4](https://doi.org/10.1016/0951-8320(95)00052-4).
33. Single, J.I.; Schmidt, J.; Denecke, J. State of Research on the Automation of HAZOP Studies. *J. Loss Prev. Process Ind.* **2019**, *62*, 103952. <https://doi.org/10.1016/j.jlp.2019.103952>.
34. Center for Chemical Process Safety. *Guidelines for Chemical Process Quantitative Risk Analysis*, 2nd ed.; AIChE/Wiley: New York, NY, USA, 2000.
35. Center for Chemical Process Safety. *Guidelines for Risk Based Process Safety*; AIChE/Wiley: Hoboken, NJ, USA, 2007.
36. Center for Chemical Process Safety. *Guidelines for Management of Change for Process Safety*; AIChE/Wiley: Hoboken, NJ, USA, 2008.

37. Center for Chemical Process Safety. *Guidelines for Auditing Process Safety Management Systems*, 2nd ed.; AIChE/Wiley: Hoboken, NJ, USA, 2011.
38. Zio, E.; Miqueles, L. Digital Twins in Safety Analysis, Risk Assessment and Emergency Management. *Reliab. Eng. Syst. Saf.* **2024**, *246*, 110040. <https://doi.org/10.1016/j.ress.2024.110040>.
39. Choi, S.; Woo, J.; Kim, J.; Lee, J.Y. Digital Twin-Based Integrated Monitoring System: Korean Application Cases. *Sensors* **2022**, *22*, 5450. <https://doi.org/10.3390/s22145450>.
40. Correa, O.C.; Mariano, J.L.S.; Amaro, E.P.; Santos, E.d.N. Process Safety Management in Oil and Gas Operating Units Through Digital Twin Platform: A Digital Approach for Safety Control and Process Intervention. In Proceedings of the Offshore Technology Conference Brasil, Rio de Janeiro, Brazil, 24–26 October 2023. <https://doi.org/10.4043/32894-ms>.
41. Rony, M.A.; Shafa, H. Cloud-Integrated Digital Twin Architectures for Real-Time Monitoring, Risk Assessment, and Safety Optimization in U.S. Energy Infrastructure. *Am. J. Interdiscip. Stud.* **2024**, *5*, 96–133. <https://doi.org/10.63125/y9m5pz24>.
42. Aljahwari, I.; Prabhu, V.V. Architecture for Dynamic Process Safety Management Using Digital Twin and Machine Learning. In *Advances in Production Management Systems. Cyber-Physical-Human Production Systems: Human-AI Collaboration and Beyond (APMS 2025)*; IFIP Advances in Information and Communication Technology; Springer: Berlin/Heidelberg, Germany, 2025; pp. 209–224. https://doi.org/10.1007/978-3-032-03515-8_15.
43. Keprate, A.; Bagalkot, N. Use of digital twins for process safety management. In *Methods in Chemical Process Safety: Methods to Assess and Manage Process Safety in Digitalized Process System*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 561–589. <https://doi.org/10.1016/bs.mcps.2022.05.001>.
44. Al Naser, I.; Rashid, A.; Bdiwi, M.; Ihlenfeldt, S. Approach for Modeling and Monitoring Hazards Towards Safety Digital Twins for Robotic Applications. In *Annals of Scientific Society for Assembly, Handling and Industrial Robotics 2023*; Springer: Cham, Switzerland, 2025. https://doi.org/10.1007/978-3-031-74010-7_22.
45. Pal, P.K.; Hens, A.; Behera, N.; Lahiri, S.K. Digital Twins: Transforming the Chemical Process Industry—A Review. *Can. J. Chem. Eng.* **2025**, *103*, 3611–3636. <https://doi.org/10.1002/cjce.25611>.
46. An, N.Y.; Yang, J.H.; Song, E.; Hwang, S.H.; Byun, H.G.; Park, S. Digital Twin-Based Hydrogen Refueling Station (HRS) Safety Model: CNN-Based Decision-Making and 3D Simulation. *Sustainability* **2024**, *16*, 9482. <https://doi.org/10.3390/su16219482>.
47. Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N.; Yin, K. A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-Based Methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311. [https://doi.org/10.1016/S0098-1354\(02\)00160-6](https://doi.org/10.1016/S0098-1354(02)00160-6).
48. Festor, P.; Nagendran, M.; Gordon, A.C.; Faisal, A.A.; Komorowski, M. Safety of Human–AI Cooperative Decision-Making within Intensive Care: A Physical Simulation Study. *PLoS Digit. Health* **2025**, *4*, e0000726. <https://doi.org/10.1371/journal.pdig.0000726>.
49. Endsley, M.R. From Here to Autonomy: Lessons Learned from Human–Automation Research. *Hum. Factors* **2017**, *59*, 5–27. <https://doi.org/10.1177/0018720816681350>.
50. Honig, C.; Desu, A.; Franklin, J. GenAI in the Classroom: Customized GPT Roleplay for Process Safety Education. *Educ. Chem. Eng.* **2024**, *49*, 55–66. <https://doi.org/10.1016/j.ece.2024.09.001>.
51. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *J. Comput. Phys.* **2019**, *378*, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
52. Brunton, S.L.; Noack, B.R.; Koumoutsakos, P. Machine Learning for Fluid Mechanics. *Annu. Rev. Fluid Mech.* **2020**, *52*, 477–508. <https://doi.org/10.1146/annurev-fluid-010719-060214>.
53. Lai, C.; Ahmed, I.; Zio, E.; Li, W.; Zhang, Y.; Yao, W.; Chen, J. A Multistage Physics-Informed Neural Network for Fault Detection in Regulating Valves of Nuclear Power Plants. *Energies* **2024**, *17*, 2647. <https://doi.org/10.3390/en17112647>.
54. Koksai, E.S.; Asrav, T.; Esenboga, E.E.; Cosgun, A.; Kusoglu, G.; Aydin, E. Physics-Informed and Data-Driven Modeling of an Industrial Wastewater Treatment Plant with Actual Validation. *Comput. Chem. Eng.* **2024**, *189*, 108801. <https://doi.org/10.1016/j.compchemeng.2024.108801>.
55. Alhajeri, M.S.; Abdullah, F.; Wu, Z.; Christofides, P.D. Physics-Informed Machine Learning Modeling for Predictive Control Using Noisy Data. *Chem. Eng. Res. Des.* **2022**, *186*, 34–49. <https://doi.org/10.1016/j.cherd.2022.07.035>.
56. Okamura, K.; Yamada, S. Empirical Evaluations of Framework for Adaptive Trust Calibration in Human–AI Cooperation. *IEEE Access* **2020**, *8*, 220335–220351. <https://doi.org/10.1109/ACCESS.2020.3042556>.
57. Single, J.I.; Schmidt, J.; Denecke, J. Knowledge Acquisition from Chemical Accident Databases Using an Ontology-Based Method and Natural Language Processing. *Saf. Sci.* **2020**, *129*, 104747. <https://doi.org/10.1016/j.ssci.2020.104747>.
58. Single, J.I.; Schmidt, J.; Denecke, J. Computer-Aided HAZOP Studies: Knowledge Representation and Algorithmic Hazard Identification. In *Proceedings of the WIT Transactions on The Built Environment*; WIT Press: Southampton, UK, 2019; Volume 189, pp. 55–66. <https://doi.org/10.2495/SAFE190061>.
59. Venkatasubramanian, V.; Vaidhyanathan, R. A Knowledge-Based Framework for Automating HAZOP Analysis. *AIChE J.* **1994**, *40*, 496–505. <https://doi.org/10.1002/aic.690400311>.

60. Single, J.I.; Schmidt, J.; Denecke, J. Ontology-Based Support for Hazard and Operability Studies. *Int. J. Saf. Secur. Eng.* **2020**, *10*, 311–319. <https://doi.org/10.18280/ijssse.100302>.
61. Zhao, C.; Bhushan, M.; Venkatasubramanian, V. PHASuite: An Automated HAZOP Analysis Tool for Chemical Processes: Part I: Knowledge Engineering Framework. *Process Saf. Environ. Prot.* **2005**, *83*, 509–532. <https://doi.org/10.1205/psep.04055>.
62. Kelly, J.; Zafar, S.A.; Heidemann, L.; Zacchi, J.V.; Espinoza, D.; Mata, N. Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-Critical Products. In *Proceedings of the 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore*; IEEE: Piscataway, NJ, USA, 2024; pp. 979–984. <https://doi.org/10.1109/cai59869.2024.00179>.
63. Zhao, Y.; Zhang, B.; Gao, D. Construction of Petrochemical Knowledge Graph Based on Deep Learning. *J. Loss Prev. Process Ind.* **2022**, *76*, 104736. <https://doi.org/10.1016/j.jlp.2022.104736>.
64. Yin, Z.; Shi, L.; Yuan, Y.; Tan, X.; Xu, S. A Study on a Knowledge Graph Construction Method of Safety Reports for Process Industries. *Processes* **2023**, *11*, 146. <https://doi.org/10.3390/pr11010146>.
65. Ekramipooya, A.; Boroushaki, M.; Rashtchian, D. Predicting Possible Recommendations Related to Causes and Consequences in the HAZOP Study Worksheet Using Natural Language Processing and Machine Learning: BERT, Clustering, and Classification. *J. Loss Prev. Process Ind.* **2024**, *89*, 105310. <https://doi.org/10.1016/j.jlp.2024.105310>.
66. Lee, J.; Park, S.; Oh, S.; Ma, B. Can Large Language Models Automate the HAZOP Process without Human Intervention? *Saf. Sci.* **2026**, *194*, 107039. <https://doi.org/10.1016/j.ssci.2025.107039>.
67. Daniel, T.; Xuan, J. Responsible Use of Generative AI in Chemical Engineering. *Digit. Chem. Eng.* **2024**, *12*, 100168. <https://doi.org/10.1016/j.dche.2024.100168>.
68. Charalampidou, S.; Zeleskidis, A.; Dokas, I.M. Hazard analysis in the era of AI: Assessing the usefulness of ChatGPT4 in STPA hazard analysis. *Saf. Sci.* **2024**, *178*, 106608. <https://doi.org/10.1016/j.ssci.2024.106608>.
69. Ricketts, J.; Guo, W.; Pelham, J.; Barry, D. Integrating an incident dataset with a question and answering language model to assist hazard identification: Comparison of an extractive and generative model. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2025**, *239*, 736–753. <https://doi.org/10.1177/1748006x241272831>.
70. Fulford, B.; Wilkinson, T.; Wylie, M.; Hankins, L. AI-Assisted HAZOP: Can Prompt Engineering Deliver a Credible HAZOP Assessment. In *Proceedings of the Safety-Critical Systems Symposium 2026, York, UK, 10–12 February 2026*.
71. Arunthavanathan, R.; Sajid, Z.; Amin, M.T.; Tian, Y.; Khan, F.; Pistikopoulos, E. Process Safety 4.0: Artificial Intelligence or Intelligence Augmentation for Safer Process Operation? *AIChE J.* **2024**, *70*, e18475. <https://doi.org/10.1002/aic.18475>.
72. Ongwae, B. An Industrial Application of a Large Language Model to Enhancing Asset Integrity and Process Safety Management. *Preprints* **2025**. <https://doi.org/10.20944/preprints202503.1172.v1>.
73. Xu, B.; Lu, D.; Gao, D.; Zhang, B. B-TBM: A Novel Deep Learning Model with Enhanced Loss Function for HAZOP Risk Classification Using Natural Language Statistical Laws. *Processes* **2024**, *12*, 2373. <https://doi.org/10.3390/pr12112373>.
74. Liu, S.; Lei, F.; Zhao, D.; Liu, Q. Abnormal Situation Management in Chemical Processes: Recent Research Progress and Future Prospects. *Processes* **2023**, *11*, 1608. <https://doi.org/10.3390/pr11061608>.
75. Meel, A.; Seider, W.D. Real-Time Risk Analysis of Safety Systems. *Comput. Chem. Eng.* **2008**, *32*, 827–840. <https://doi.org/10.1016/j.compchemeng.2007.03.006>.
76. Cai, B.; Liu, Y.; Fan, Q.; Zhang, Y.; Liu, Z.; Yu, S.; Ji, R. Multi-Source Information Fusion Based Fault Diagnosis of Ground-Source Heat Pump Using Bayesian Network. *Appl. Energy* **2014**, *114*, 1–9. <https://doi.org/10.1016/j.apenergy.2013.09.043>.
77. Parasuraman, R.; Manzey, D.H. Complacency and Bias in Human Use of Automation: An Attentional Integration. *Hum. Factors* **2010**, *52*, 381–410. <https://doi.org/10.1177/0018720810376055>.
78. Khan, F.; Rathnayaka, S.; Ahmed, S. Methods and Models in Process Safety and Risk Management: Past, Present and Future. *Process Saf. Environ. Prot.* **2015**, *98*, 116–147. <https://doi.org/10.1016/j.psep.2015.07.005>.
79. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. <https://doi.org/10.1038/s42254-021-00314-5>.
80. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *ACM Comput. Surv.* **2022**, *55*, 1–37. <https://doi.org/10.1145/3514228>.
81. von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Pfrommer, J.; Pick, A.; Ramamurthy, R.; et al. Informed Machine Learning: A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 614–633. <https://doi.org/10.1109/TKDE.2021.3079836>.
82. Jang, K.; Pilario, K.E.S.; Lee, N.; Moon, I.; Na, J. Explainable Artificial Intelligence for Fault Diagnosis of Industrial Processes. *IEEE Trans. Ind. Inform.* **2025**, *21*, 4–11. <https://doi.org/10.1109/TII.2023.3240601>.
83. Rajesh Alias Harinarayan, R.; Mercy Shalinie, S. XFDDC: EXplainable Fault Detection Diagnosis and Correction Framework for Chemical Process Systems. *Process Saf. Environ. Prot.* **2022**, *165*, 463–474. <https://doi.org/10.1016/j.psep.2022.07.019>.
84. Yuan, Y.; Chaffart, D.; Wu, T.; Zhu, J. Transparency: The Missing Link to Boosting AI Transformations in Chemical Engineering. *Engineering* **2024**, *39*, 45–60. <https://doi.org/10.1016/j.eng.2023.11.024>.
85. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.

86. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
87. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Diaz-Rodriguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. <https://doi.org/10.1016/j.inffus.2023.101805>.
88. Wen, H.; Amin, M.T.; Khan, F.; Ahmed, S.; Imtiaz, S.; Pistikopoulos, E. Assessment of Situation Awareness Conflict Risk between Human and AI in Process System Operation. *Ind. Eng. Chem. Res.* **2023**, *62*, 4028–4038. <https://doi.org/10.1021/acs.iecr.2c04310>.
89. Endsley, M.R. Automation and situation awareness. In *Automation and Human Performance: Theory and Applications*; Parasuraman, R., Mouloua, M., Eds.; CRC Press: Boca Raton, FL, USA, 2018.
90. Lebiere, C.; Blaha, L.M.; Fallon, C.K.; Jefferson, B. Adaptive Cognitive Mechanisms to Maintain Calibrated Trust and Reliance in Automation. *Front. Robot. AI* **2021**, *8*, 652776. <https://doi.org/10.3389/frobt.2021.652776>.
91. Stewart, K.; Buchan, G.; Rahnama, R.; Abdullah, M.; Radu, A. Enhancing Process Safety Management: A Novel Application of GPT-Based Statistical AI Models for Driving Improvement in Process Safety Barrier Health. In *Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference (ADIPEC)*, Abu Dhabi, UAE, 4–7 November 2024. <https://doi.org/10.2118/223053-ms>.
92. Rashed, M.; Mishra, A. Safety Studies Using Artificial Intelligence. In *Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference (ADIPEC)*, Abu Dhabi, UAE, 3–6 November 2025; p. SPE-230047-MS. <https://doi.org/10.2118/230047-ms>.
93. Winfield, A.F.T.; Jirotko, M. Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence Systems. *Philos. Trans. R. Soc. A* **2018**, *376*, 20180085. <https://doi.org/10.1098/rsta.2018.0085>.
94. Perez-Cerrolaza, J.; Abella, J.; Borg, M.; Donzella, C.; Cerquides, J.; Cazorla, F.J.; Englund, C.; Tauber, M.; Nikolakopoulos, G.; Flores, J.L. Artificial Intelligence for Safety-Critical Systems in Industrial and Transportation Domains: A Survey. *ACM Comput. Surv.* **2024**, *56*, 1–40. <https://doi.org/10.1145/3626314>.
95. Qin, S.J. Survey on Data-Driven Industrial Process Monitoring and Diagnosis. *Annu. Rev. Control* **2012**, *36*, 220–234. <https://doi.org/10.1016/j.arcontrol.2012.09.004>.
96. Jiao, Z.; Hu, P.; Xu, H.; Wang, Q. Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. *ACS Chem. Health Saf.* **2020**, *27*, 316–334. <https://doi.org/10.1021/acs.chas.0c00075>.
97. Amin, M.T.; Imtiaz, S.; Khan, F. Process System Fault Detection and Diagnosis Using a Hybrid Technique. *Chem. Eng. Sci.* **2018**, *189*, 191–211. <https://doi.org/10.1016/j.ces.2018.05.045>.
98. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
99. National Aeronautics and Space Administration. *Technology Readiness Level Definitions*; Technical report; NASA: Washington, DC, USA, 2012.
100. Ekramipooya, A.; Boroushaki, M.; Rashtchian, D. Application of Natural Language Processing and Machine Learning in Prediction of Deviations in the HAZOP Study Worksheet: A Comparison of Classifiers. *Process Saf. Environ. Prot.* **2023**, *176*, 65–73. <https://doi.org/10.1016/j.psep.2023.06.004>.
101. Lee, J.; Oh, S.; Ma, B.; Park, S. Advancing HAZOP Automation: A Comparative Evaluation of Multimodal Large Language Models. *SSRN Preprint* **2025**. <https://doi.org/10.2139/ssrn.5084592>.
102. Noufal, M. A Unified AI-ChemE Framework for Physics-Constrained Modeling and Digital-Twin Control: Case Studies in Fluidized Bed Heat Transfer and PEM Fuel Cells. *ChemRxiv Preprint* **2025**. <https://doi.org/10.26434/chemrxiv-2025-s1chw>.
103. Zhou, Z.; Wang, X.; Yan, Y.; Mijiddorj, L.; Ding, Y.; Beringer, T.; Khiabani, P.M.; Jentner, W.G.; Hu, X.M.; Wang, C.; et al. AIMNET: An IoT-Empowered Digital Twin for Continuous Gas Emission Monitoring and Early Hazard Detection. *arXiv* **2025**, arXiv:2512.06148. <https://doi.org/10.48550/arXiv.2512.06148>.
104. Berdouzi, F.; Villemur, C.; Olivier-Maget, N.; Gabas, N. Dynamic Simulation for Risk Analysis: Application to an Exothermic Reaction. *Process Saf. Environ. Prot.* **2018**, *113*, 149–163. <https://doi.org/10.1016/j.psep.2017.09.019>.
105. Wang, Z.; Zhang, B.; Gao, D. A Novel Knowledge Graph Development for Industry Design: A Case Study on Indirect Coal Liquefaction Process. *Comput. Ind.* **2022**, *139*, 103647. <https://doi.org/10.1016/j.compind.2022.103647>.
106. Thiebes, S.; Lins, S.; Sunyaev, A. Trustworthy Artificial Intelligence. *Electron. Mark.* **2021**, *31*, 447–464. <https://doi.org/10.1007/s12525-020-00441-4>.
107. Pasman, H.J.; Reniers, G. Past, Present and Future of Quantitative Risk Assessment (QRA) and the Incentive It Obtained from Land-Use Planning (LUP). *J. Loss Prev. Process Ind.* **2014**, *28*, 2–9. <https://doi.org/10.1016/j.jlpi.2013.03.004>.
108. Hermann, E.; Hermann, G.; Tremblay, J.C. Ethical Artificial Intelligence in Chemical Research and Development: A Dual Advantage for Sustainability. *Sci. Eng. Ethics* **2021**, *27*, 45. <https://doi.org/10.1007/s11948-021-00325-6>.

109. Huang, X.; Kroening, D.; Ruan, W.; Sharp, J.; Sun, Y.; Thamo, E.; Wu, M.; Yi, X. A Survey of Safety and Trustworthiness of Deep Neural Networks: Verification, Testing, Adversarial Attack and Defence, and Interpretability. *Comput. Sci. Rev.* **2020**, *37*, 100270. <https://doi.org/10.1016/j.cosrev.2020.100270>.
110. International Electrotechnical Commission. *IEC 61508: Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems*, 2nd ed.; Technical report; International Electrotechnical Commission: Geneva, Switzerland, 2010.
111. Braband, J.; Schäbe, H. On Safety Assessment of Artificial Intelligence. *Dependability* **2020**, *20*, 25–34. <https://doi.org/10.21683/1729-2646-2020-20-4-25-34>.
112. UL Standards. *UL 4600: Standard for Safety for the Evaluation of Autonomous Products*; Technical report; UL Standards: Evanston, IL, USA, 2020.
113. Koopman, P.; Wagner, M. Autonomous Vehicle Safety: An Interdisciplinary Challenge. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 90–96. <https://doi.org/10.1109/MITS.2016.2583491>.

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.