Process Safety and Environmental Protection 222 (2026) 109780

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.journals.elsevier.com/process-safety-and-environmental-protection

How a digital twin for process safety will differ from a production twin

Kamran Gholamizadeh [a,b], Stewart W. Behie [d], Hans J. Pasman [c], Ming Yang [e,*]

[a] Department of Systems Engineering, École de Technologie Supérieure, University of Quebec, Montréal, Canada
[b] Risk Analytics AI (RAAI), AZ, USA
[c] Emeritus Professor, Chemical Risk Management, Delft University of Technology, Delft, the Netherlands
[d] Retired, Chief Process Safety, Occidental Oil and Gas, Corporate Headquarters, Visiting Professor Industrial Engineering, Texas A&M University Corpus Christi, Ocean Drive, TX, USA
[e] Safety and Security Science Section, Faculty of Technology, Policy and Management, TU Delft, Delft, the Netherlands

ARTICLE INFO

Keywords:
Digital twin
Process safety
Early warning
Dynamic risk assessment
Safety barrier health
Anomaly detection
Bow-Tie
Validation
Loss prevention

ABSTRACT

To alleviate complexity, digital twins (DTs) have moved rapidly from a concept to a widely promoted manufacturing support technology across various types of industry, including the process industries. The literature offers abundant demonstrations of DTs for manufacturing used to monitor the functioning of the process assets to provide information on anomalies. So, it will be great for production optimization and training operators to control upsets. With this, safety is at stake, yet the conditions under which DTs can be trusted to inform safety-critical decisions remain poorly defined. No structured account is available of what distinguishes a safety-qualified digital twin from a sophisticated process safety control dashboard. This paper asserts that such a safety twin addressing that gap requires five functional types of twins for their primary safety role. Further, validation evidence is required before deployment and autonomy may be granted to them, and the governance obligations they carry. The typology is mapped onto seven elements of process safety management (PSM). For its realization, a six-layer reference architecture is proposed that spans physical data acquisition, process modelling and plant degradation, predictive analytics and fault diagnosis, dynamic risk assessment and safety intelligence, human-centred decision support, and governance, together with a tiered validation hierarchy anchored in functional safety practice. Finally, two reproducible Python case demonstrations ground the framework: 1) a multivariate anomaly-detection for a simulated continuous distillation unit, which detected an abrupt condenser fault within 8 min and two developing faults within operationally useful horizons at a false-alarm rate of one event per day, and 2) a digitised Bow-Tie for a pressurised hydrocarbon reactor, in which barrier health degradation raised the top-event likelihood proxy by 21.2% over 55 months and in which a sensitivity analysis isolated the single most risk-critical barrier. The paper closes with a critical appraisal of data dependency, uncertainty communication, and the organisational gap between demonstration and deployment.

1. Introduction

Major accidents in the process industries continue to occur. Besides unidentified hazards, particularly at start-ups, and human and organizational failures, this is due to a familiar combination of degraded safeguards, unnoticed process drift, and decisions based on information that no longer describes the true state of the plant, e.g., underestimated risk of failure of a temporary repair. Post-incident investigations repeatedly show that the evidence needed to interrupt the accident sequence existed somewhere in the facility, distributed across historian records, inspection findings, proof-test logs, and operator observations, but was never assembled into a coherent, current picture of risk (Kletz, Amyotte, 2019; Pasman, 2015). The instruments of conventional process safety management, including hazard and operability studies, layer of protection analysis, and quantitative risk assessment, are performed at intervals of months to years, and the risk picture they produce begins to age the moment it is documented (Khan et al., 2015).

Digital twin technology promises to close precisely this gap. A digital twin, in the sense introduced by (Grieves, Vickers, 2016) and consolidated in later surveys (Rasheed et al., 2020; Singh et al., 2024; Tao et al., 2018), maintains a virtual representation of a physical asset that is synchronised with live operational data and can therefore answer questions about the present and projected state of the system rather than about its state at the last formal review. Within the process industries,

* Corresponding author.
E-mail addresses: kamran.gholami-zadeh.1@ens.etsmtl.ca (K. Gholamizadeh), M.Yang-1@tudelft.nl (M. Yang).

https://doi.org/10.1016/j.psep.2026.109780
Received 20 August 2026; Received in revised form 30 September 2026; Accepted 21 October 2026
Available online 22 October 2026
0957-5820/© 2026 The Authors. Published by Elsevier Ltd on behalf of Institution of Chemical Engineers. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the concept has been applied to plant monitoring, control optimisation, predictive maintenance, and operator training (Pal et al., 2025; Perno et al., 2020; Peterson et al., 2025), and a growing subset of studies addresses safety-relevant functions such as fault diagnosis, dynamic risk assessment, and emergency simulation (Agnusdei et al., 2021; Keprate, Bagalkot, 2022; Zio, Miqueles, 2024).

The promise, however, has outrun the discipline. A digital twin whose outputs influence a permit-to-work authorisation, an inspection deferral, or an emergency response decision occupies a position in the safety decision chain comparable to that of a safety instrumented function, yet it is typically developed, validated, and governed to no recognised standard at all. The literature provides no structured answer to three questions that any process safety engineer must ask before relying on such a system: what class of safety function is this twin's reliability and trustworthiness; what validation evidence should exist before its outputs are trusted for that function; and what governance obligations follow from that trust. In the absence of answers, organisations either over-trust systems that are technically impressive but unvalidated for safety use, or decline to deploy systems that could deliver genuine risk reduction.

This paper develops a framework that answers those three questions and demonstrates its analytical content computationally. Three contributions are made. First, a functional typology is proposed that classifies process safety digital twins into five types according to their primary safety function, and attaches to each type a required validation level, an acceptable autonomy boundary, and a governance tier. The typology is mapped onto seven core elements of the process safety management (PSM), enabling practitioners to locate each twin type within existing management systems. Second, a six-layer reference architecture is specified for safety-critical digital twins, covering the full chain from field instrumentation to governance, and is complemented by a tiered validation hierarchy grounded in the philosophy of IEC 61508 and IEC 61511 functional safety practice (Commission, 2003; Leander et al., 2019). Third, two reproducible computational demonstrations, one at the anomaly-detection tier and one at the risk-and-barrier tier, show that the analytical capabilities the framework describes are achievable with current methods, while also exposing the data-quality and uncertainty constraints that any honest deployment assessment must confront.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature and distills the research gaps the framework addresses. Section 3 outlines the approach used to construct the framework. Section 4 presents the functional typology and its PSM mapping. Section 5 develops the six-layer reference architecture and the validation hierarchy. Section 6 reports the two computational demonstrations. Section 7 discusses implementation constraints and research needs, and Section 8 concludes. The paper is a synopsis of a comprehensive chapter prepared by the authors for the fifth edition of Lees' Loss Prevention in the Process Industries; the chapter provides extended treatment, two additional demonstrations, and a practical eight-step deployment guideline.

2. Related work and research gaps

Research on digital twins in the process industries has expanded along several fronts that are relevant to safety but were rarely designed for it. This section synthesises that body of work thematically and identifies the specific gaps that motivate the framework of Sections 4 and 5. The scope of the synthesis is the process industries in the wide sense, including chemicals, petrochemicals, oil and gas, and adjacent high-hazard sectors whose safety practice rests on the same barrier and PSM foundations. Studies from construction, manufacturing, and nuclear domains are cited where they contribute concepts, notably reference-model thinking and risk-informed operational decision making, that transfer to process applications; sector-specific results that do not transfer are excluded.

The synthesis followed a structured protocol. Scopus and Web of Science were searched as the primary databases, supplemented by IEEE Xplore and ScienceDirect for engineering and computational sources. The search combined the term "digital twin" with process safety descriptors, including "process safety," "risk assessment," "safety barrier," "fault detection," "anomaly detection," "validation," "functional safety," and "emergency management." Coverage spanned from the consolidation of the digital twin concept in the mid-2010s to the most recent literature available at the time of writing. Studies were included when they concerned digital twins, or their component methods, in the process industries or in adjacent high-hazard domains whose safety practice rests on the same barrier and process-safety-management foundations, and when they bore on a safety function, a risk or barrier method, a validation or governance question, or an operator-facing decision. Studies were excluded when they addressed production optimisation with no safety relevance, or when their results were specific to a sector whose conclusions do not transfer to process safety. The retained records were organised into the eight themes listed in Section 3, and representative rather than exhaustive coverage was sought within each theme, consistent with the aim of motivating a framework rather than producing a bibliometric census.

2.1. Conceptual foundations and reviews

The conceptual lineage of the digital twin runs from the mirrored-systems idea of (Grieves, Vickers, 2016) through the industrial state-of-the-art surveys of (Tao et al., 2018) and (Rasheed et al., 2020) to process-industry-specific reviews (Mane et al., 2024; Pal et al., 2025; Perno et al., 2020). Computational and numerical foundations for process engineering twins have recently been consolidated by (Peterson et al., 2025). Across these reviews a consistent observation recurs: classification schemes organise twins by technology (physics-based versus data-driven) or by connectivity (batch-updated versus continuously synchronised), not by the function the twin performs in a safety management system. A practitioner seeking guidance on validation or governance with rules and standards for a safety application finds none in these taxonomies.

2.2. Safety-oriented applications

A second stream addresses safety functions directly. Bibliometric and systematic reviews confirm accelerating interest in digital twins for safety management (Agnusdei et al., 2021). (Zio, Miqueles, 2024) provide the most complete account of twins in safety analysis, risk assessment, and emergency management. Application studies span offshore risk-informed decision support (Murray et al., 2019; Vessel, 2020), oil and gas process safety platforms (Al Ghazal Abdullah et al., 2022; Correa et al., 2023; Mohd Sahak et al., 2022), petrochemical monitoring and fault diagnosis architectures (Hu et al., 2021), operator training simulators (Kaarlela et al., 2020; Zabarin, Shilin, 2023), ergonomics and occupational safety (Jia et al., 2024; Park et al., 2023), hazardous materials handling (Li et al., 2022), and nuclear operational decision making (Muhlheim et al., 2025). Reference models for operator risk prevention in process plants have been proposed (Bevilacqua et al., 2020), and the Bow-Tie itself has been examined from a data perspective as a candidate twin structure (Singh et al., 2024). The embedding of a digital twin directly within the safety-risk-management process has also been demonstrated, coupling process simulation with operational risk control. These studies demonstrate feasibility in individual niches but do not supply a unifying architecture, and almost none state what validation evidence preceded safety-relevant use.

2.3. Dynamic risk assessment and barrier thinking

Dynamic risk analysis predates the digital twin label and supplies much of its safety-relevant analytical machinery. The evolution of dynamic methods in the chemical and petroleum industries is documented

by (Paltrinieri, Khan, 2016), and the broader methodological trajectory of process safety modelling by (Khan et al., 2015). Barrier theory, in particular, the definitions and performance concepts of (Sklet, 2006), provides the vocabulary through which degraded safeguards can be represented quantitatively. Machine learning has been positioned as a means of assessing risk from operational data, together with clear warnings about its limits for rare events (Paltrinieri et al., 2019). Recent work has begun to couple data-driven and physics-based inference with dynamic reliability and risk analysis. (Wu et al., 2025) embed physical degradation knowledge in a physics-informed dynamic Bayesian network to predict the time-dependent, multi-state reliability of a subsea wellhead sealing system, showing how domain constraints can discipline a temporal model where failure data are scarce. Along the data-driven axis, (Wu et al., 2024) apply machine learning to the intelligent identification of drilling lost-circulation risk, which demonstrates both the reach of learned models over operational data and their dependence on representative training records. On the barrier side, the definitions and performance concepts of (Sklet, 2006) are complemented by the ARAMIS treatment of (De Dianous, Fievez, 2006), which links barrier classification and performance evaluation to Bow-Tie representation, and by dynamic-coupling analyses in severe-service equipment, where interacting failure mechanisms erode protection in ways that a static barrier list cannot capture. What is missing is the connective tissue: a specification of how barrier states, degradation models, and dynamic risk metrics should be arranged into a live system whose outputs are trustworthy enough for safety decisions.

2.4. Validation, trustworthiness, and functional safety

The functional safety standards (Commission, 2003; Leander et al., 2019) define a mature philosophy for qualifying systems whose failure endangers people, but they were written for deterministic hardware and software functions, not for continuously learning statistical models. Early work on classifying digital twins for functional safety use and on tool qualification has begun to appear (Mitov, 2025), and the machine learning community has developed uncertainty concepts, including the aleatoric and epistemic distinction (Hüllermeier, 2021), conformal prediction for distribution-free guarantees, and post-hoc explanation methods (Arrieta et al., 2020). A recent consensus study on machine learning for safety-critical applications stresses that these ingredients have not yet been assembled into deployable qualification pathways (Gholamizadeh, 2026; Gholamizadeh, Zarei, Gualtieri, et al., 2025). Model-quality manifestos from the broader modelling community carry the same message (Saltelli et al., 2020).

2.5. Gaps addressed by this paper

Four gaps emerge from this synthesis. First, no classification of digital twins exists that is organised by safety function and linked to validation and governance requirements. Second, no reference architecture specifies the layers a safety-critical twin must contain and the obligations each layer carries. Third, no tiered validation scheme connects the functional safety tradition to data-driven twin components. Fourth, published demonstrations rarely disclose reproducible implementations against which detection performance, false-alarm behaviour, and risk-metric dynamics can be examined. Table 1 substantiates these gaps by comparing representative digital twin frameworks across the attributes that a safety-oriented framework must satisfy. None of the surveyed frameworks combines classification by safety function with a consequence-scaled validation obligation, explicit barrier modelling, a defined autonomy boundary, and lifecycle governance. Sections 4–6 of this paper address these four gaps in turn.

3. Framework development approach

The framework was constructed in three stages. In the first stage, a structured review of the digital twin literature across the process industries and adjacent safety-critical domains was performed, organised into eight themes (conceptual foundations; architecture and data integration; validation and trustworthiness; dynamic risk assessment; monitoring and fault diagnosis; control and decision support; operator training and human factors; and emergency management). The synthesis of that review yielded the four gaps summarised in Section 2.5 and a set of design requirements that any safety-oriented framework must satisfy: it must accommodate the rarity of safety events, treat uncertainty conservatively, represent safety barriers explicitly, and preserve human authority over safety-critical actions.

In the second stage, the DT typology and architecture were developed deductively from those requirements and iteratively tested against the application studies of Section 2.2: each published safety application

Table 1
Comparison of representative digital twin frameworks with the framework proposed in this study, across attributes relevant to process safety assurance. Entries indicate whether each attribute is addressed (Yes), partially addressed (Partial), or not addressed (No).

Framework	Organizing basis	Safety-function classification	Consequence-scaled validation	Barriers as first-class objects	Autonomy boundary	Governance and reproducible demonstration
(Grieves, Vickers, 2016)	Conceptual mirroring of a physical asset	No	No	No	No	No
(Tao et al., 2018)	Industrial twin components and maturity	No	No	No	No	Partial
(Rasheed et al., 2020)	Enabling technologies, modelling perspective	No	No	No	No	No
(Hu et al., 2021)	Reference architecture for petrochemical monitoring and fault diagnosis	No	No	No	No	Partial
(Agnusdei et al., 2021)	Bibliometric review of twins for safety management	Partial	No	No	No	No
(Agrell et al., 2021)	Decision-theoretic probabilistic twin	No	Partial	No	No	No
(Zio, Miqueles, 2024)	Application areas in safety, risk, and emergency management	Partial	Partial	Partial	Partial	No
(Singh et al., 2024)	Data-perspective representation of a Bow-Tie	No	No	Yes	No	No
This study	Safety function performed by the twin	Yes (five types)	Yes (tiered hierarchy)	Yes	Yes (defined per type)	Yes (Layer 6; two demonstrations)

It should be noted that Zio, Miqueles, (2024), which is assigned the qualification 'partial' to four of the five entries, has reviewed 89 references rather than a specific DT architecture. The studies reviewed covered DT-enabled possibilities, various manufacturing and construction activities, and only one on process plants. So, the 'partial' is related to different studies and does not need to be relevant in view of the complexity of process plants.

was assigned to a candidate type and layer set, and the classification was refined until every surveyed application could be located without ambiguity. The validation hierarchy was derived by transposing the graded-rigour logic of IEC 61508 and IEC 61511, in which the required evidence scales with the safety significance of the function, onto the five twin types.

In the third stage, two computational demonstrations were implemented in Python to verify that the analytical capabilities attributed to the lower and middle tiers of the typology are realisable with current methods, and to quantify their behaviour under controlled fault injection (**see Appendix File**). Synthetic simulation was chosen deliberately: it provides labelled ground truth for fault onset times, which operational data cannot supply, and it makes the demonstrations fully reproducible. The corresponding limitation, that synthetic performance does not transfer directly to operating plants, is examined in Section 7. Fig. 1 presents the framework as a single end-to-end workflow. Qualified field data enter at acquisition (Layer 1) and pass through digital-model construction and analytics (Layers 2 and 3) into dynamic risk analysis (Layer 4), whose barrier-availability and top-event-likelihood outputs are delivered to operators through the decision-support layer (Layer 5). Two features distinguish this workflow from a production-optimisation one. First, a tiered validation gate precedes decision support, so that no output enters a safety-relevant decision until it carries the evidence its tier requires (Section 5.2). Second, a dynamic-updating loop returns proof-test, inspection, condition-monitoring, and near-miss evidence to the models, keeping the twin current as barriers age and conditions change, while a governance layer spans the entire life cycle. The workflow also situates the demonstrations of Section 6: Demonstration A exercises the acquisition, analytics, and risk stages of a Type I function, and Demonstration B exercises the risk stage of a Type II function.

4. A functional typology of process safety digital twins

The typology proposed here classifies process safety digital twins by the safety function they perform rather than by the technology they employ. Five types are identified. They are not maturity stages and they are not mutually exclusive; a single facility may operate several types simultaneously, each serving different users and supporting different decisions. What distinguishes the types is the consequence of being wrong: as the twins' outputs move closer to safety-critical action, the evidence demanded before deployment, and the governance imposed during operation must both increase. Table 2 summarizes the five types of process safety digital twins together with their validation, autonomy, and governance attributes, and Fig. 2 depicts the typology graphically.

Before the types are introduced, it is worth stating what makes a *safety twin* categorically different from the production-optimisation twins that dominate the literature, because the typology's validation and governance attributes follow from these differences. A production twin is judged by the average quality of its predictions over the conditions it routinely encounters; a safety twin is judged by its behaviour in the rare, severe conditions it may encounter once in its lifetime, which means the two systems face inverted data economies. A wrong production twin loses money in a bounded and recoverable way; a safety twin that is wrong can license an inspection deferral or suppress an alert whose consequences are neither bounded nor recoverable, which is why the asymmetry of error costs must be built into thresholds and uncertainty handling rather than left to interpretation. A production twin can treat safeguards as constraints on the optimisation; a safety twin must treat safeguards as first-class modelled objects whose availability, degradation, and demand history are tracked explicitly (Sklet, 2006). And a production twin may legitimately act autonomously within its economic mandate, whereas a safety twin operates inside a decision

Table 2
Functional typology of process safety digital twins: safety function, validation requirement, autonomy boundary, and governance tier.

Type	Primary safety function	Required validation focus	Autonomy boundary	Governance tier
I	Continuous monitoring and early anomaly detection	Detection rate, latency, and false-alarm rate against normal plus fault data	Advisory only; alerts and trends	Basic: model owner, periodic performance review
II	Dynamic risk, barrier health, and remaining-useful-life estimation	Calibration of risk quantities; traceable degradation bases; uncertainty bounds	Advisory; informs inspection and maintenance prioritisation	Intermediate: documented model basis, revalidation triggers
III	Abnormal-situation management training	Behavioural fidelity of process response; avoidance of negative training transfer	No plant connection; simulation only	Intermediate: scenario library control, fidelity audits
IV	Guided protective intervention	Speed and accuracy of consequence projection; usability under time pressure	Advisory to emergency command	High: drill-based verification, availability assurance
V	Guided or semi-autonomous protective intervention	Functional-safety-grade qualification of any action pathway	Bounded actions under explicit human authorisation	Highest: SIS-equivalent lifecycle management

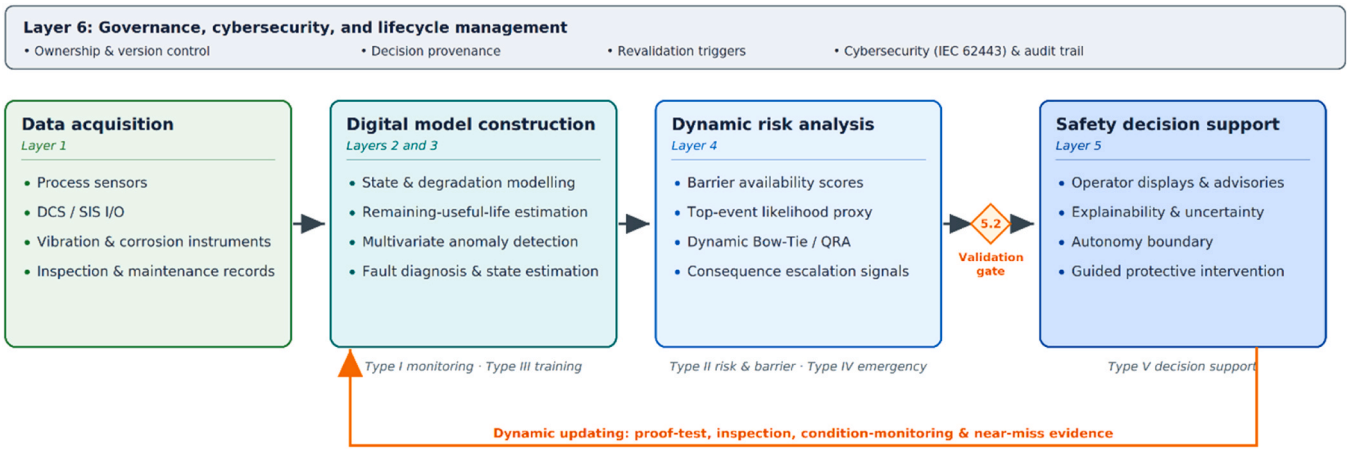


Fig. 1. End-to-end workflow of a process safety digital twin.

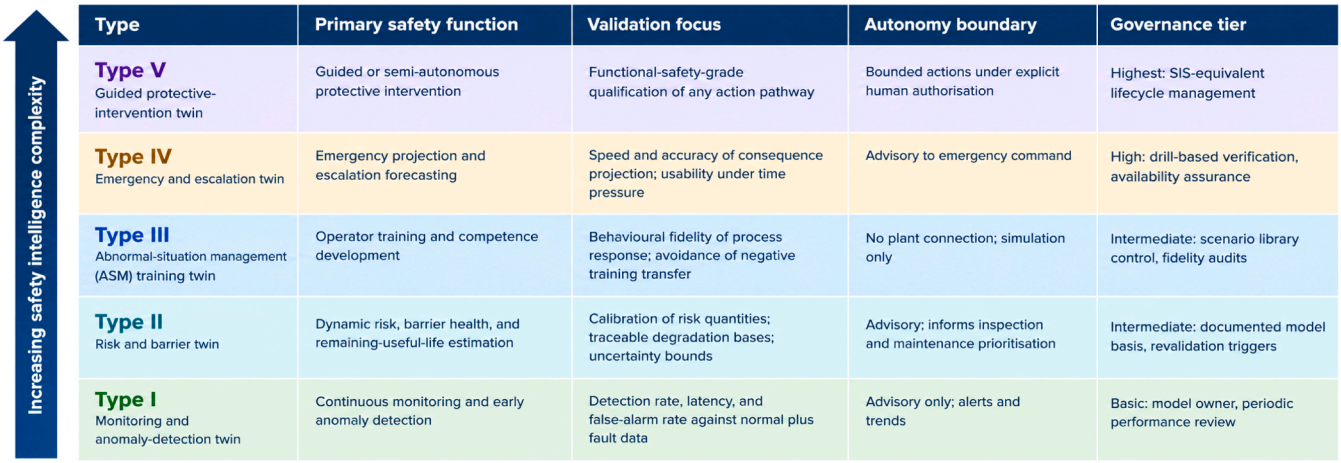


Fig. 2. The five functional types of process safety digital twins, ordered by increasing safety intelligence complexity, showing the primary safety function, validation focus, autonomy boundary, and governance tier of each type. ASM, abnormal-situation management; SIS, safety instrumented system.

chain in which human authority over safety-critical action is itself a protective barrier that automation must not silently erode. The five types below encode these differences at increasing levels of consequence.

Each type also resolves a specific limitation of general-purpose digital twin frameworks. The monitoring and anomaly-detection twin (Type I) makes explicit the alarm-management and data-validation duties that production monitoring twins leave implicit. The risk and barrier twin (Type II) represents safety barriers as tracked objects with quantified availability and degradation, rather than as constraints on an optimisation. Isolation from live systems and behavioural fidelity, seldom required of general simulators, are imposed on the abnormal-situation management training twin (Type III). Benchmarking against recognised consequence models, not demanded of production twins, is required of the emergency and escalation twin (Type IV). The guided protective-intervention twin (Type V) bounds permissible autonomy and preserves the human-authority boundary that existing autonomous twins leave unstated.

4.1. *The five types of process safety digital twins*

Type I, the monitoring and anomaly-detection twin, performs continuous surveillance of process and equipment state and raises alerts when behaviour departs from a learned or modelled normal envelope. Typical implementations of fault detection and diagnosis (FDD) rest on multivariate statistical process monitoring, autoencoder, recurrent-network detectors, a recently proposed self-gated hierarchical transformer encoder, or physics-informed residual models. Its outputs are advisory trend indicators and anomaly scores; it takes no action. Validation must establish detection rate, detection latency, and false-alarm rate against a dataset containing both normal variability and representative fault signatures.

Type II, the risk and barrier twin, translates equipment condition and process state into explicit risk quantities: barrier availability estimates, degradation indices, remaining-useful-life distributions, and top-event likelihood indicators. Because its outputs feed inspection prioritisation, possible maintenance deferral, and management-of-change judgements, its validation must extend beyond detection statistics to the calibration of the risk quantities themselves, and its degradation models must be traceable to recognised engineering bases such as published S-N curves or reliability data collections.

Type III, the abnormal-situation management (ASM) training twin, provides a high-fidelity simulated environment in which operators rehearse abnormal-situation recognition and response. Its validation criterion is behavioural fidelity: the simulated process must respond to interventions in a way that builds correct rather than misleading operator expectations. Negative training transfer from an inaccurate simulator is a genuine safety hazard, which is why this type carries a validation burden that purely advisory tools do not.

Type IV, the emergency and escalation twin, projects the development of an unfolding abnormal event, including consequence footprints and escalation pathways, in support of emergency response decisions. It operates under the most severe time pressure of the five types and must therefore be validated for speed as well as accuracy, and its interface must be designed for use under acute cognitive load (Endsley, 2017).

Type V, the guided protective-intervention twin, closes part of the loop by recommending or initiating protective interventions subject to defined human authorisation. It is the only type whose outputs may translate directly into plant actions, and it therefore inherits the full weight of functional safety expectations: its permissible autonomy must be bounded explicitly, any action pathway that could constitute a safety function must be qualified in accordance with IEC 61511 (2016), and recent work coupling large language model agents to fault handling in process plants illustrates how quickly capability in this tier is advancing ahead of qualification practice (Gill et al., 2025).

In vocabulary more familiar from safety instrumented system (SIS) practice, Type I is safety-supporting, in that it informs without itself entering the protective chain; Types II to IV are safety-critical to varying degrees, in that their outputs materially shape protective decisions without executing them; and Type V is safety-qualified, in that its outputs may execute a protective action and must therefore meet functional-safety-grade qualification. Type V accordingly carries the highest governance obligation of the five types. This safety-instrumented-system vocabulary is offered as a bridge to established terminology, not as an additional twin type.

4.2. *Positioning within the process safety management*

A typology is practically useful only if each type can be located within the management system. Table 3 maps the five types onto seven core PSM elements drawn from established risk-based process safety practice (CCPS, 2010): hazard identification, mechanical integrity, management of change, permit-to-work, training and competence, emergency preparedness, and incident learning. Two observations follow from the mapping. First, in a plant life cycle, no single type is sufficient; a facility pursuing comprehensive coverage needs to operate the whole portfolio of types. Second, the mapping exposes department ownership. For example, a Type II twin that informs mechanical integrity decisions must be owned within the mechanical integrity element, with its outputs formally connected to that element's decision

Table 3
Mapping of the five twin types onto seven core PSM elements (● primary contribution; ○ supporting contribution).

PSM element	Type I	Type II	Type III	Type IV	Type V
Hazard identification	●	○		○	
Mechanical integrity	○	●			○
Management of change	○	●			○
Permit-to-work	○	●			○
Training and competence			●	○	
Emergency preparedness		○	○	●	○
Incident and near-miss learning	●	○	○	○	

procedures, or its insights will remain analytically correct and operationally inert.

The mapping also clarifies what integration means in practice for the three elements where Type II contributes most. In mechanical integrity, the twin's degradation indices and remaining-useful-life estimates should enter the inspection-planning procedure as a formal input alongside the risk-based inspection basis, with a documented rule for how twin-derived urgency modifies the scheduled interval. In management of change, the twin serves twice: as a screening input that flags when accumulated barrier degradation makes a proposed change less tolerable than the original design intent assumed, and as a subject of change control itself, since retraining a model or repointing a data source is a modification to a safety-relevant system and must pass through the same discipline as a physical modification. In permit-to-work, a live barrier-status view gives the issuing authority what a static isolation register cannot: the current, quantified availability of the safeguards the permit relies on at the moment of authorisation.

5. A six-layer reference architecture and tiered validation hierarchy

Published twin implementations differ so widely in scope and internal structure that their safety-relevant properties cannot be compared (Azangoo et al., 2022; Hu et al., 2021; Qamsane et al., 2022). A reference architecture resolves this by naming the six functional information-processing layers that an integral safety-critical twin, comprising the five partial twin types, must contain. For each layer, it specifies the output owed to the layer above and the validation obligation the layer carries. The six-layered architecture proposed here, is shown in Fig. 3. Layers 1–5 form the analytical chain from field data to human decision; Layer 6 is a cross-cutting governance layer that spans the entire system throughout its life. Data are fed from and in return fed to the physical plant. The system will require an assurance level.

5.1. The analytical layers

- Layer 1, physical data acquisition, delivers time-stamped, quality-flagged measurements from field instrumentation, laboratory analyses, inspection records, and maintenance logs. Its safety-specific obligation is data qualification: every variable that feeds a safety-relevant computation must have a documented reliability basis, a known failure signature, and a defined behaviour when the measurement itself is degraded.
- Layer 2, process and degradation modelling, maintains the physics of the plant: conservation, thermodynamic and kinetic process models, damage-accumulation and ageing models, and the hybrid physics-plus-data formulations that dominate current practice, including physics-informed machine learning (Lei et al., 2018; Raissi et al., 2019).
- Layer 3, analytics and fault diagnosis, houses the statistical and machine-learning machinery for state estimation, anomaly

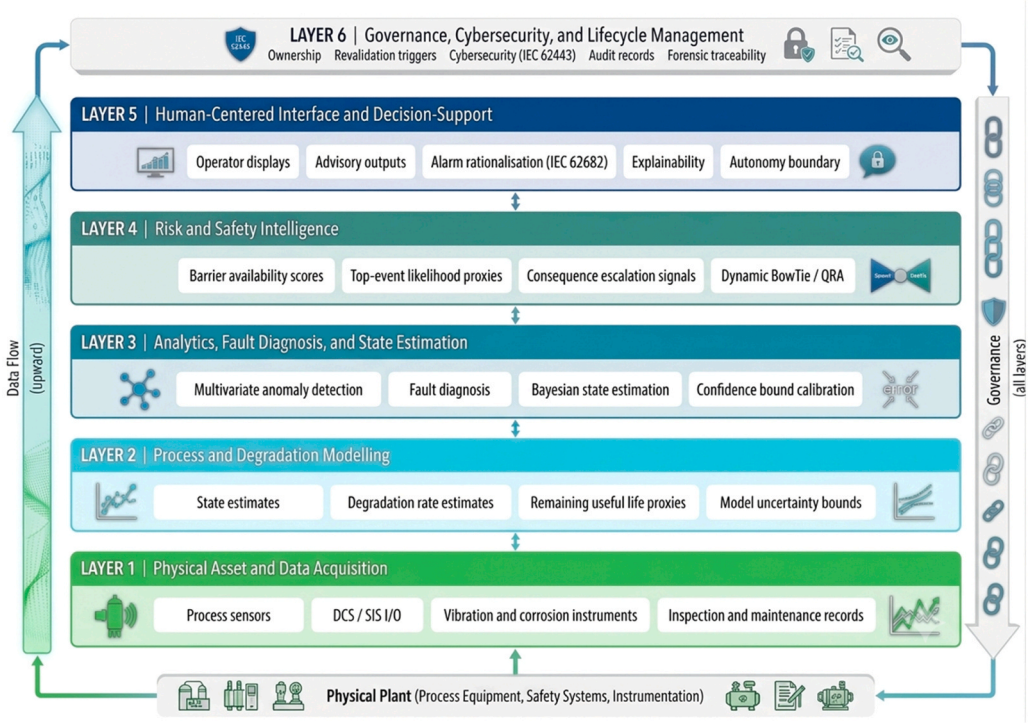


Fig. 3. Six-layer reference architecture for safety-critical process digital twins. Layers 1–5 form the analytical chain from field data to human decision; Layer 6 governs the system across its operational life; data stream from and to the physical plant. DCS, distributed control system; SIS, safety instrumented system; QRA, quantitative risk assessment; IEC 62443, industrial communication networks and system security; IEC 62682, management of alarm systems for the process industries.

detection, and fault isolation; multivariate monitoring statistics and their control limits live here.

- Layer 4, risk and safety intelligence, is the layer that most sharply distinguishes a safety twin from a production twin. It converts the outputs of Layers 2 and 3 into quantities with direct safety meaning: barrier availability scores, top-event likelihood indicators, escalation probabilities, and composite risk scores whose thresholds are anchored to the facility's risk criteria. Dynamic quantitative risk assessment and living Bow-Tie constructions are Layer 4 implementations (Paltrinieri, Khan, 2016; Singh et al., 2024).
- Layer 5, human-centred decision support, delivers the Layer 4 risk quantities, namely the barrier availability scores, top-event likelihood indicators, escalation probabilities, and composite risk scores, to operators and engineers in a form that supports correct action under load: alarm integration consistent with recognised alarm-management practice (Commission, 2014; EEMUA, 2013), presentation of trend and threshold context, explicit display of uncertainty, and on-demand explanation of what is driving any alert (Arrieta et al., 2020).
- Layer 6, governance, assigns a formally accountable owner, defines revalidation triggers and suspension criteria, and imposes cybersecurity controls appropriate to a safety-connected system (Cindrić et al., 2025; Wang et al., 2024). It also maintains decision provenance and version control: every result or recommendation the twin produces is traceable to the specific model version, training-data snapshot, parameter set, and human input that generated it, and a record links each output to the model version in force when it was produced. This provenance underpins the forensic traceability that incident investigation will one day demand, and it becomes more important precisely because the framework requires models to be updated regularly.

The layered decomposition earns its place through three properties that a monolithic design cannot offer. First, it localises validation: each layer owes a defined product to the layer above, so evidence can be assembled layer by layer, and a change confined to one layer, such as retraining an anomaly detector, triggers revalidation of that layer and its downstream consumers rather than of the entire system. Second, it localises failure: a corrupted sensor is a Layer 1 event whose quality flag should prevent silent propagation into Layer 4 risk indicators, and an out-of-domain model excursion is a Layer 3 event that Layer 5 must surface to the user as reduced confidence rather than conceal behind an unchanged display. Third, it makes responsibilities assignable: instrument engineers own Layer 1 qualification, process modellers own Layer 2, data scientists own Layer 3, process safety engineers own Layer 4, human factors specialists own Layer 5, and a named senior owner carries Layer 6, so that the recurring failure mode of digitalisation programmes, capability without accountability, is designed out rather than managed after the fact. The architecture is compatible with, and deliberately more prescriptive than, the general five-dimension twin models in the literature, because safety-critical use demands prescriptions where general-purpose frameworks may legitimately remain agnostic (Rasheed et al., 2020; Tao et al., 2018).

5.2. The tiered validation hierarchy

The validation hierarchy connects the architecture to the typology. Its organising principle is borrowed from functional safety: evidence must scale with consequence (Häring, 2021). The hierarchy is a proposal by the authors, not a codified standard. It transposes the graded-rigour philosophy of functional safety (IEC 61508 and IEC 61511) (Bell, 2006; Smith, Simpson, 2020) onto the five twin types and reflects the authors' engineering judgement and practitioner experience, and it is offered as an interim, facility-adaptable template intended to structure validation practice until standardisation bodies define a recognised qualification pathway (Sections 7.3 and 8). Three complementary evidence categories are defined. 1) Empirical validation demonstrates statistical performance against datasets that contain representative fault behaviour, reporting detection rate, latency, false-alarm rate, and, for Type II functions, the calibration of risk quantities. 2) Formal verification examines the model itself: boundary-condition behaviour, physical-consistency checks, code review, and explicit documentation of the domain within which the model's outputs are meaningful. 3) Independent expert validation subjects the whole construction to review by qualified process safety practitioners who did not build it, testing whether the scenario coverage, the assumptions, and the intended use are jointly defensible.

Table 4 assigns minimum evidence expectations from each category to each twin type. The hierarchy is deliberately demanding at the upper tiers: a Type V twin whose action pathway can constitute a safety function must meet SIS-equivalent lifecycle expectations, and no accumulation of in-distribution accuracy statistics can substitute for that.

The hierarchy also requires that each validation package name the uncertainties it does and does not address, using the now-standard distinction between aleatoric variability, the irreducible randomness of the process, and epistemic uncertainty, the reducible ignorance embedded in model structure, parameters, and training coverage (Hüllermeier, 2021). The distinction is operational, not philosophical: aleatoric uncertainty is managed through thresholds and persistence logic, whereas epistemic uncertainty is managed through model qualification, domain-of-validity documentation, and revalidation triggers, and a package that quantifies the first while remaining silent on the second gives an illusion of completeness that the upper tiers of the typology cannot tolerate.

Three design principles cut across all layers and types and deserve an explicit statement. The rare-event principle acknowledges that the events the twin exists to prevent are almost absent from operational data, so validation datasets must be constructed, through fault injection, physics-based simulation, or transfer from analogous units, rather than merely collected (Paltrinieri, Khan, 2016). The conservative-uncertainty principle requires that, where the twin is uncertain, its outputs err toward protective action, and that uncertainty be carried visibly through the analytical chain rather than collapsed into confident point estimates (Hüllermeier, 2021; Saltelli et al., 2020). The human-authority principal reserves safety-critical action for humans unless a specific action pathway has been qualified to functional safety standards; the twin advises, and people decide.

6. Computational demonstrations

Two demonstrations, both implemented in Python with fixed random seeds and released as reproducible scripts, ground the

Table 4
Tiered validation hierarchy: minimum evidence expectations by twin type.

Type	Empirical validation	Formal verification	Independent expert validation
I	Detection rate, latency, false-alarm rate on normal plus fault data	Training-window integrity; control-limit basis documented	Review of fault-scenario coverage
II	Calibration of risk and RUL quantities; uncertainty bounds verified	Degradation model traceable to a recognised engineering basis	Review of model basis and intended decisions
III	Behavioural fidelity across the scenario library	Process-response verification against reference simulations	Instructor and human-factors review
IV	Projection accuracy and computation-time budget under drill conditions	Consequence-model verification; availability assurance	Emergency-command exercise review
V	All of the above plus intervention-pathway reliability	Functional-safety-grade verification of action logic	SIS-equivalent independent assessment

framework in executable form. Demonstration A exercises the Type I tier and Layers 1, 3, and 4 of the architecture through multivariate anomaly detection on a simulated continuous distillation unit. Demonstration B exercises the Type II tier and Layer 4 through a digitised Bow-Tie whose barrier health scores evolve over a five-year horizon. The demonstrations are illustrative rather than plant-validated; their purpose is to quantify what the framework's lower and middle tiers can deliver under controlled conditions and to expose the constraints that Section 7 examines.

6.1. Demonstration A: multivariate anomaly detection in a continuous distillation unit

Demonstration A illustrates the validation evidence that a Type I monitoring twin must supply, and it is presented as an instance of that tier rather than as a standalone anomaly-detection study.

A six-tray continuous distillation column was simulated at one-minute resolution over 6000 minutes, with eight monitored variables: six tray temperatures, the column pressure drop, and the reflux ratio. Realistic cross-variable correlation was created by driving all trays with a shared first-order autoregressive disturbance, representing common-cause variation in feed and reboiler duty, with tray-specific loadings and additive local noise. Three safety-relevant fault modes were injected at known times. Incipient flooding (minutes 1500–2300) was represented by a ramped rise in pressure drop, growing erratic behaviour of the lower-tray temperatures, and a compensating operator increase of reflux ratio. Condenser malfunction (minutes 3000–3700) was represented by an abrupt sustained step in the two uppermost tray temperatures with a moderate pressure rise. Progressive loss of tray efficiency (from minute 4400) was represented by a slow flattening of the temperature profile, the kind of drift signature that defeats univariate alarms.

The three fault modes were chosen to span the detection-difficulty spectrum rather than to flatter the method. The condenser step is the easy case: an abrupt, sustained break in the learned correlation structure that any competent multivariate scheme should catch quickly. The flooding case is intermediate: a developing hydraulic constraint whose signature grows over hours and is partially masked by the operator's compensating reflux action, which itself is part of the injected scenario because real operators respond to developing upsets and their responses reshape the data the twin sees. The efficiency-loss case is deliberately hard: a slow, structured drift in which no individual variable behaves alarmingly at any moment, so that detection must come from the cumulative departure of the joint profile from normality. A simulated detection pipeline that handles only the first case has demonstrated little; the value of the exercise lies in quantifying how latency degrades across the spectrum and in showing which architectural element, the static limits or the adaptive drift channel, carries each case.

The detection pipeline follows the Layer 3 to Layer 4 chain of the architecture. A principal component analysis model was fitted to a 1,200-minute clean baseline after standardisation, retaining the components explaining 90% of variance (three components, 91.1% explained). Hotelling's T-squared statistic and the squared prediction error (SPE) were computed for every sample, with control limits set at the 99th percentile of their baseline distributions (10.44 and 2.48, respectively). To recover sensitivity to slow drift, an exponentially weighted moving average (EWMA) of the T-squared statistic (smoothing constant 0.06) was tracked against an adaptive limit set four baseline standard deviations above the baseline mean. Layer 4 then fused the three indicators into a composite risk score on a 0–100 scale (Eq. 4), normalising each indicator by its own control limit, capping it at three times that limit, and weighting the Hotelling, squared-prediction-error, and moving-average contributions at 0.45, 0.35, and 0.20. The score was smoothed over a centred 15-minute window, and an alarm was confirmed only after ten consecutive minutes above a threshold of 55, a persistence filter that suppresses transient spikes. The detector statistics are defined as follows, for a standardised observation $\mathbf{z}^t$ with score vector $\mathbf{t}_t = \mathbf{P}^\top \mathbf{z}_t$ on the retained loadings $\mathbf{P}$ and eigenvalue matrix $\mathbf{\Lambda}$ as Eq. (1) to Eq. (4).

$$T_t^2 = \mathbf{t}_t^\top \, \mathbf{\Lambda}^{-1} \, \mathbf{t}_t \tag{1}$$

$$\mathrm{SPE}_t = \left\| \mathbf{z}_t - \mathbf{P}\mathbf{P}^\top \mathbf{z}_t \right\|^2 \tag{2}$$

$$g_t = \lambda \, T_t^2 + (1-\lambda) \, g_{t-1}, \lambda = 0.06 \tag{3}$$

$$\begin{aligned} u(x,c) &= \mathrm{clip}(x/c, 0, 3)/3, S_t \\ &= 100\left[0.45u\left(T_t^2, c_{T^2}\right) + 0.35u(\mathrm{SPE}_t, c_{\mathrm{SPE}}) + 0.20u\left(g_t, c_g\right)\right] \end{aligned} \tag{4}$$

The control limits $c\,T^2$ and $c\,SPE$ are the 99th percentiles of the respective baseline distributions (10.44 and 2.48), and the adaptive limit $c\,g$ for the moving average is placed four baseline standard deviations above the baseline mean. Full parameter values and the data-generation equations are given in Appendix A (Tables A1 and A2).

Fig. 4 shows the resulting monitoring statistics across the full simulation, and Fig. 5 shows the composite risk score with confirmed alarms. The two statistics respond in physically interpretable ways: the SPE, which measures departure from the learned correlation structure, reacts violently to the condenser step change, while the T-squared statistic tracks the growing magnitude of the flooding deviation. The drifting efficiency fault raises both statistics only gradually, and it is the EWMA channel that first accumulates the evidence into a confirmed alarm, vindicating the adaptive-threshold design choice.

Table 5 summarises detection performance against the known fault onsets. The abrupt condenser fault was confirmed 8 min after onset. The developing flooding fault was confirmed after 179 min, roughly three hours into an event whose hydraulic escalation timescale comfortably exceeds that horizon. The slow efficiency fault was confirmed after 436 min, again well within the timescale over which intervention remains effective for a degradation mechanism of this kind. Over the 48.3 h of normal operation, the persistence-filtered pipeline produced false alarms at 0.041 events per hour, approximately one per day, a rate compatible with recognised alarm-management expectations (Commission, 2014). The spread of latencies across the three faults carries a methodological message: detection latency is a joint property of the fault signature and the monitoring logic, so a twin validated only against abrupt faults offers no evidence about the slow precursors that often matter most.

6.2. Demonstration B: dynamic barrier health monitoring with a digitised Bow-Tie

Demonstration B illustrates the additional validation a Type II risk and barrier twin must satisfy, beyond the detection statistics of Type I, and it is likewise presented as an instance of its tier rather than as a standalone barrier-monitoring study.

The second demonstration addresses the question a static Bow-Tie cannot answer: how much protection does the plant actually have today? The top event is loss of primary containment from a pressurised flammable hydrocarbon reactor. Three threats feed the knot: process overpressure (0.12 events per year), external corrosion (0.04 per year), and instrument failure leading to overfill (0.08 per year). Five preventive barriers guard the threat lines: a pressure relief valve and a high-pressure instrumented trip on the overpressure path (nominal probabilities of failure on demand of 0.10 and 0.01), corrosion monitoring with inhibitor injection and an inspection-and-maintenance programme on the corrosion path (0.20 and 0.15), and a high-level instrumented trip on the overfill path (0.10). Five mitigative barriers shape the consequence side: fire and gas detection (0.05), emergency isolation (0.10), deluge (0.15), emergency depressurisation (0.10), and evacuation (0.05), arranged along three consequence pathways culminating in a vapour cloud explosion.

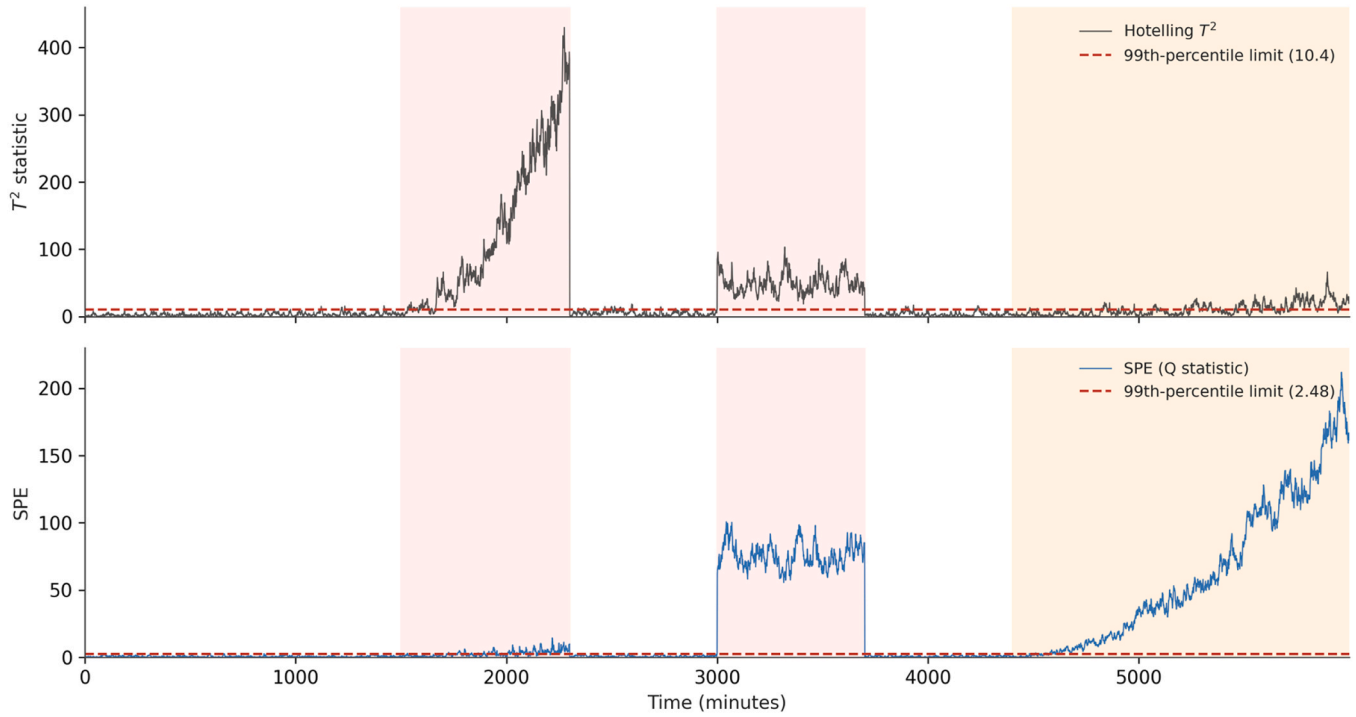


Fig. 4. Layer 3 monitoring statistics for Demonstration A: Hotelling's T-squared (upper) and SPE (lower) with 99th-percentile control limits; injected fault windows shaded.

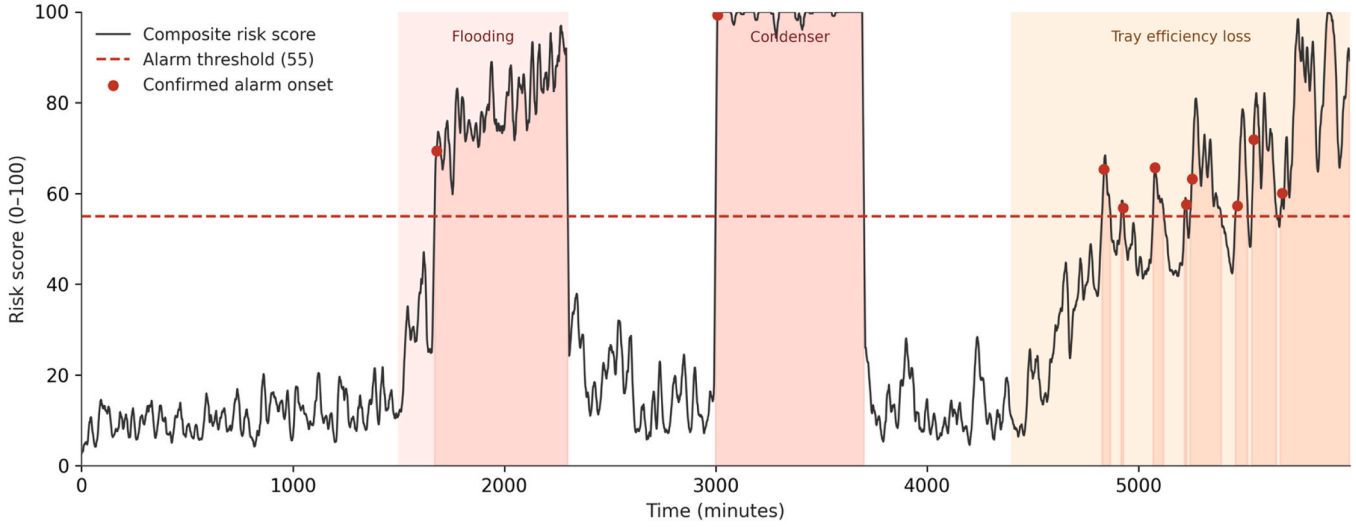


Fig. 5. Layer 4 composite risk score for Demonstration A with alarm threshold and persistence-confirmed alarm onsets.

Table 5
Demonstration A detection performance against known fault onsets.

Fault mode	Onset (min)	Confirmed (min)	Latency	Dominant detection channel
Incipient flooding	1500	1679	179 min (~3.0 h)	Joint T-squared and SPE rise with pressure-drop ramp
Condenser malfunction	3000	3008	8 min	Immediate SPE spike (correlation-structure break)
Tray efficiency loss	4400	4836	436 min (~7.3 h)	EWMA drift channel ahead of static limits

False-alarm rate over normal operation: 0.041 events per hour (approximately one event per day over 48.3 normal-operation hours).

The choice of the Bow-Tie as the risk representation for this tier is deliberate. The quantity the twin must expose is barrier availability, and the Bow-Tie is the one classical representation that places preventive and mitigative barriers on a single causal structure organised around a single top event, a representation whose use for barrier-performance evaluation is well established (Aliabadi, Gholamizadeh, 2021; Gholamizadeh, Zarei, Yazdi, Ramezanifar, et al., 2024; Mohammadfam et al., 2020), so that degradation on any line reads directly as a change in protection. Fault Tree Analysis and Event Tree Analysis decompose, respectively, the causes and the consequences of that top event with greater combinational detail, but each addresses only one side of the knot and, like the Bow-Tie, represents static logic; neither foregrounds barrier condition as the object of interest. Dynamic Bayesian Networks are, in principle, better suited to dynamic risk assessment, because they

represent temporal dependency, common-cause coupling, and Bayesian updating as evidence arrives, and a mature Type II deployment would be expected to adopt Dynamic Bayesian Network or probabilistic-digital-twin propagation as the technical substrate beneath the barrier-health quantities, as Section 7.2 discusses. Their cost is transparency: they require conditional probability tables and data that brownfield facilities rarely hold, and their inference is less directly auditable by the process safety engineers who must act on the output, which is a Layer 5 concern. The Bow-Tie was therefore selected here for interpretability and auditability in a demonstration whose purpose is to show that barrier condition can be propagated to a top-event likelihood in a form its intended users already trust, consistent with recent work representing the Bow-Tie from a digital-twin data perspective (Singh et al., 2024). The living Bow-Tie mitigates the principal limitation of the static method by making barrier health time-varying, although it does not, in this form, capture the temporal dependencies that a Dynamic Bayesian Network would.

The digitisation assigns every barrier a time-varying health score H between 0 and 1, driven by gradual ageing between proof tests, saw-tooth restoration at each test, and two injected discrete events: a partial failure of the high-pressure trip between months 32 and 38, and a deluge system failure between months 45 and 51. At each monthly step, the effective probability of failure on demand is obtained by dividing the nominal value by the health score; the top-event likelihood (TEL) proxy is recomputed as the sum over threats of initiating frequency times the product of effective preventive-barrier failure probabilities, and conditional consequence probabilities are recomputed along each mitigative pathway. The formulation is deliberately transparent rather than maximally sophisticated: every quantity in it can be audited by a process safety engineer, which is itself a Layer 5 design decision. Formally, each barrier i carries a health score $H_i(t) in [0,1]$, and its effective probability of failure on demand is computed as Eq. (5).

$$\mathrm{PFD}_{\mathrm{eff},i}(t) = \min\left(1, \frac{\mathrm{PFD}_{\mathrm{nom},i}}{\max(H_i(t), 0.01)}\right) \tag{5}$$

The top-event likelihood proxy aggregates the initiating threat frequencies with the series combination of the preventive barriers on each threat line as Eq. (6).

$$\mathrm{TEL}(t) = \sum_{j \in \mathcal{F}} f_j \prod_{i \in \mathrm{prev}(j)} \mathrm{PFD}_{\mathrm{eff},i}(t) \tag{6}$$

Where f_j is the initiating frequency of threat j and prev(j) is the set of preventive barriers guarding it. The reported change is the relative variation of this quantity between commissioning and the snapshot as Eq. (7).

$$\Delta\mathrm{TEL}\%(t) = \frac{\mathrm{TEL}(t) - \mathrm{TEL}(0)}{\mathrm{TEL}(0)} \times 100 \tag{7}$$

The indicator is a relative, barrier-driven proxy rather than an absolute frequency. Because the initiating threat frequencies f_j are held constant, a change in TEL(t) reflects only the degradation of preventive-barrier availability, so the increase from 0.00932 to 0.01130 per year (ΔTEL = +21.2%) measures the erosion of preventive protection alone and should be read as a proportional change in the modelled top-event likelihood contributed by the barriers, not as a calibrated plant frequency. The consequence side is treated separately: conditional outcome probabilities are recomputed as the product of the effective mitigative-barrier probabilities along each pathway. Three modelling assumptions underlie the formulation: conditional independence of barriers on a given threat line, series (product) combination along that line, and multiplicative coupling between the nominal probability of failure on demand and the health score, with the health score floored at 0.01 and the effective probability capped at unity. Threat frequencies and nominal probabilities of failure on demand are illustrative values representative of typical process-industry ranges; all parameters and their bases are listed in Appendix A (Table A2).

The health-score functions encode three degradation archetypes that mechanical integrity practice will recognise. Test-restored ageing, applied to the relief valve, the corrosion-monitoring system, and the fire and gas detection, produces the characteristic saw-tooth in which condition declines between proof tests and partially recovers at each test; the recovery is deliberately incomplete, reflecting the reality that testing restores demonstrated function without reversing accumulated physical wear. Slow monotonic ageing, applied to the remaining continuously maintained barriers, represents gradual capability loss bounded below by a maintenance floor. Discrete failure with delayed repair, applied to the instrumented trip and the deluge system, represents the revealed faults that dominate real barrier unavailability: the health score collapses at the failure, climbs through the repair window, and resumes ageing from the restored level. In an operating deployment, these functions would be replaced by estimators fed from proof-test results, maintenance closures, condition monitoring, and demand histories; the simulation's contribution is to show that once such estimators exist, the propagation from barrier condition to top-event likelihood is arithmetically simple and fully auditable. The parametric forms of the three archetypes and their per-barrier constants are given in Appendix A (Table A2), and the top-event likelihood computation is defined there for full reproducibility.

Fig. 6 presents the living Bow-Tie at the month-55 snapshot, with each barrier coloured by health tier. The relief valve has degraded to H = 0.56 through fouling accumulation between annual tests, tripling its effective failure probability; corrosion monitoring and fire and gas detection sit in the degraded band; the remaining seven barriers are healthy. Fig. 7 traces the health trajectories of five representative barriers and the normalised TEL proxy across the full 60 months: the trip failure at month 32 produces a visible step in the TEL curve, the repair partially recovers it, and the deluge failure at month 45 renews the climb. Table 6 gives the complete numerical snapshot. The TEL proxy has risen from 0.00932 to 0.01130 events per year, an increase of 21.2%, without any change in the initiating threat frequencies: the plant has become measurably less protected while every conventional indicator of its risk documentation remained unchanged.

A sensitivity analysis at the snapshot, shown in Fig. 8, halves each barrier's health score in turn and records the resulting change in the TEL proxy (preventive side) or in the vapour cloud explosion pathway probability (mitigative side). The result is starkly concentrated. Halving the health of the high-level trip would raise the TEL proxy by 96.4%, an order of magnitude more than any other preventive barrier, because it stands alone on its threat line while every other threat is guarded by two barriers in series. The corrosion-path barriers contribute 22.3% each; the overpressure-path barriers, protected by redundancy, contribute only 2.5% each. On the mitigative side, the series structure makes the four explosion-pathway barriers formally equivalent at 100% each. The operational implication is direct: uniform allocation of inspection attention across barriers is systematically suboptimal, and the twin converts that qualitative intuition into a continuously updated quantitative ranking.

6.3. Lessons across the demonstrations

A final lesson concerns what the demonstrations, set against the full typology, say about the ordering of validation demand. The two executed examples illustrate the lower tiers directly: Demonstration A, a Type I monitoring twin, is validated by the evidence Table 5 prescribes for its tier, namely detection rate, latency, and false-alarm rate against representative faults; Demonstration B, a Type II risk and barrier twin, adds calibration of its risk quantities against an auditable engineering basis and an exposed sensitivity structure. The upper tiers, characterised here rather than implemented, extend the same logic: a Type III abnormal-situation management training twin would be verified by behavioural fidelity across its scenario library and by instructor and

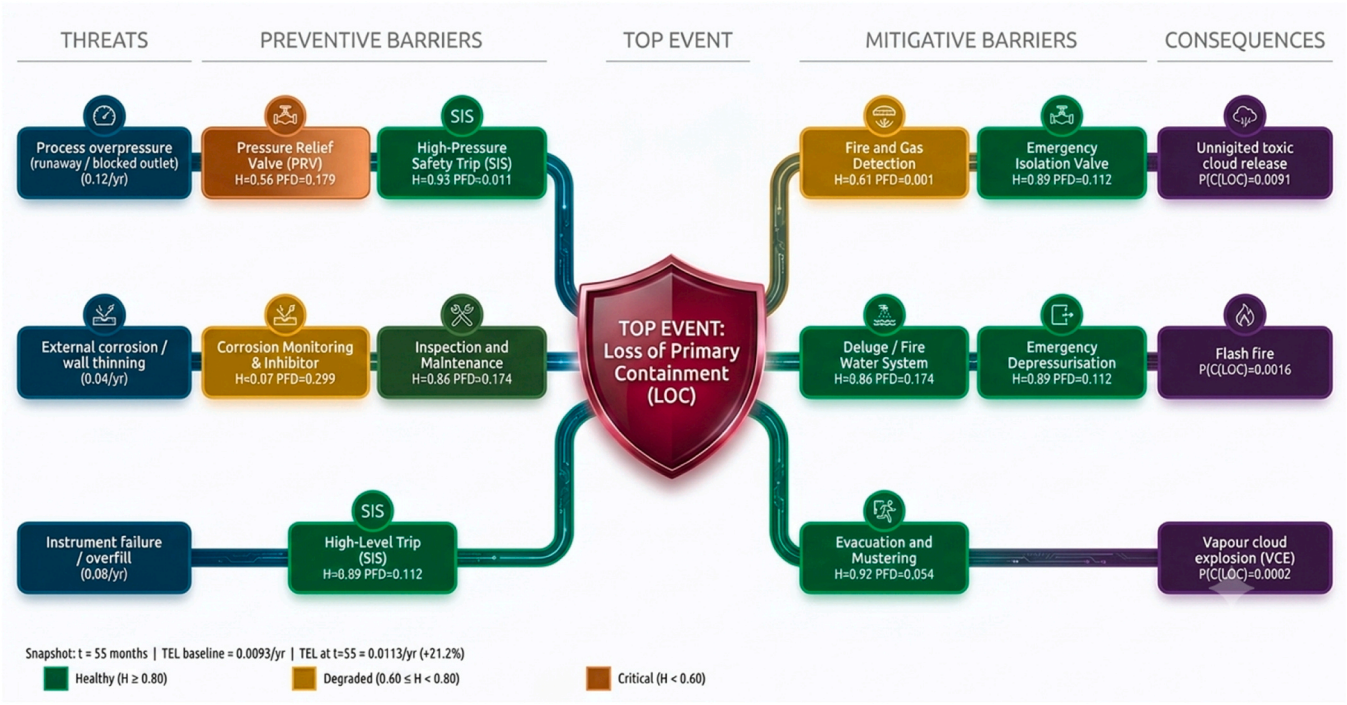


Fig. 6. Living Bow-Tie for Demonstration B at the month-55 snapshot. Barrier boxes show current health score and effective probability of failure on demand, coloured by health tier.

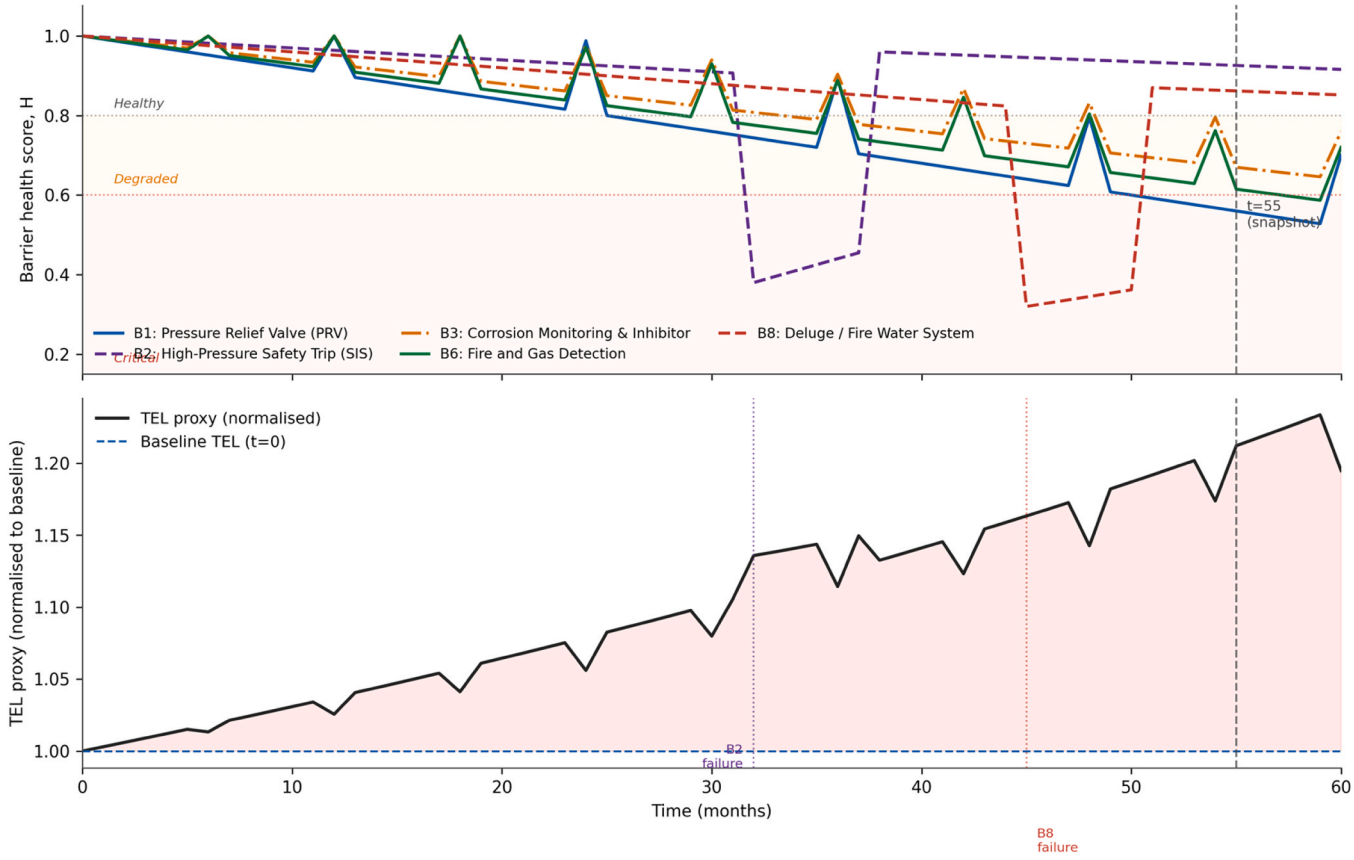


Fig. 7. Barrier health trajectories for five representative barriers (upper) and the normalised top-event likelihood proxy (lower) over the 60-month simulation, with injected barrier failures marked.

human-factors review, since negative training transfer is the hazard specific to that tier; a Type IV emergency twin would be verified for

Table 6
Demonstration B barrier state at the month-55 snapshot and resulting top-event likelihood.

Barrier	Function	Nominal PFD	H at t = 55	Effective PFD	Status
B1	Pressure relief valve	0.10	0.56	0.179	Critical
B2	High-pressure trip (SIS)	0.01	0.93	0.011	Healthy
B3	Corrosion monitoring and inhibitor	0.20	0.67	0.299	Degraded
B4	Inspection and maintenance	0.15	0.86	0.174	Healthy
B5	High-level trip (SIS)	0.10	0.89	0.112	Healthy
B6	Fire and gas detection	0.05	0.61	0.081	Degraded
B7	Emergency isolation valve	0.10	0.89	0.112	Healthy
B8	Deluge / fire water	0.15	0.86	0.174	Healthy
B9	Emergency depressurisation	0.10	0.89	0.112	Healthy
B10	Evacuation and mustering	0.05	0.92	0.054	Healthy

Top-event likelihood proxy: 0.00932 per year at commissioning; 0.01130 per year at month 55 (+21.2%).

projection accuracy and computation-time budget under drill conditions and through emergency-command exercise review; and a Type V guided protective-intervention twin would require, in addition, functional-safety-grade verification of its action logic and an independent, SIS-equivalent assessment. This gradient is the typology working as intended: the framework's claim is not that all twins are equally attainable, but that the evidence each requires is a defined function of the safety consequence of its outputs (Table 5). Two demonstrations were therefore developed in depth, to make the lower tiers reproducible, rather than five in outline.

Read together, the demonstrations verify the framework's central claims and expose its central constraints. On the affirmative side, both pipelines run in real time on commodity hardware, both produce outputs whose safety meaning is direct rather than inferred, and both quantify behaviour that static methods cannot see: a correlation-structure break eight minutes after it occurs, and a 21.2% erosion of protection that no document revision would have recorded. On the cautionary side, three constraints recur. Every quantity in both pipelines is bounded by input data quality; the PCA (principal component analysis) control limits inherit any contamination of the training baseline, and the barrier health scores are only as current as the proof-test and maintenance records that feed them. Uncertainty transforms at every layer; the 12-minute smoothing that suppresses false alarms in Demonstration A purchases that benefit with detection latency, a trade that must be documented, not discovered. And interpretability is a design property, not an afterthought: the Bow-Tie demonstration owes much of its practical credibility to the fact that its structure coincides with a representation process safety engineers already trust, whereas the multivariate statistics of Demonstration A require a deliberate Layer 5 translation before an operator can act on them.

For both demonstrations, the parameter-determination methods, sources, assumptions, intermediate computations, and validation are stated explicitly and gathered for reference. Demonstration A derives its detector settings from the fault-free baseline (standardisation statistics, retained variance, and 99th-percentile control limits), with the generative coefficients fixed in the released code; its assumptions are the common-cause and additive-noise structure of Section 6.1; its intermediate computations are Eq. (1) to Eq. (4); and its validation rests on labelled fault onsets, reported as detection latency and false-alarm rate. Demonstration B draws its threat frequencies and nominal probabilities of failure on demand from illustrative values representative of typical process-industry ranges; its assumptions are conditional independence of barriers on a threat line, series combination along that line, and multiplicative coupling between nominal probability and health score (Section 6.2); its intermediate computations are Eq. (5) to Eq. (7); and its validation rests on the auditability of every quantity against its stated basis and on the exposed sensitivity structure. All parameters with their bases are listed in Appendix A (Tables A1 and A2), and the two validation modes correspond to the empirical-validation and formal-verification categories of the tiered hierarchy (Table 5).

7. Discussion

7.1. The rarity problem and the limits of learned normality

Process safety is defined by events that operational data almost never contain. A model that learns normality from a historian record learns, with high fidelity, the statistics of a plant behaving well; it acquires no direct knowledge of the precursor signatures it exists to detect. The demonstrations of Section 6 sidestepped this problem by injecting faults with known onset times, and that is precisely the point: for real deployments, validation datasets must be manufactured through fault injection, physics-based simulation, or transfer from analogous units, because they cannot be harvested (Acuña, 2026; Paltrinieri et al., 2019). Even then, an irreducible epistemic residue remains, since a twin

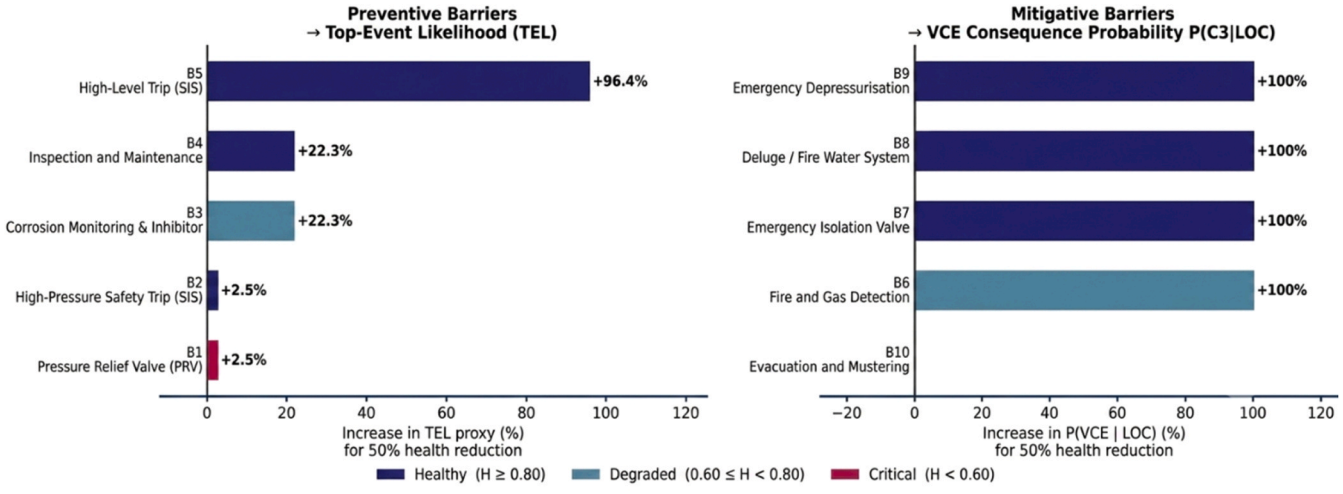


Fig. 8. Barrier sensitivity analysis at month 55: change in top-event likelihood (preventive barriers, left) and in vapour cloud explosion pathway probability (mitigative barriers, right) for a 50% reduction in each barrier's health score.

validated against constructed faults has not been validated against the specific failure that will eventually occur. The governance consequence is that near-miss reports, inspection findings, events considered as precursors, and incident investigations must be treated as continuing validation evidence, formally fed back into the twin throughout its life, rather than as inputs consumed once at commissioning. For this, after sufficient event numbers are registered over time, a procedure, such as the hierarchical Bayesian analysis, can be followed to predict the rare catastrophic event estimated probability (Gholamizadeh, Pasman, 2026; Gholamizadeh, Zarei, & BahooToroody, 2025; Gholamizadeh, Zarei, Gualtieri, et al., 2025; Gholamizadeh, Zarei, et al., 2024a; Mohammadfam et al., 2022; Zarei et al., 2022).

7.2. False assurance and the duty of calibrated uncertainty

A poorly calibrated twin that presents confident outputs is more dangerous than no twin, because it licenses decisions at a precision the evidence does not support. The distinction between in-distribution accuracy and out-of-distribution reliability is decisive here: a classifier can be near-perfect on held-out data from its own training distribution and silently unreliable the day the feed composition changes. Concept drift produces no alarm of its own. Mitigations exist and belong in the architecture from the outset: calibration monitoring that compares stated confidence with realised accuracy on recent data, conformal methods that wrap predictions in distribution-free validity guarantees (Agnusdei et al., 2021; Lee et al., 2025), out-of-distribution detection that tells the user when the model is being asked about conditions it has never seen, and revalidation triggers wired into the governance layer. These are Layer 4 and Layer 6 obligations under the present framework, not optional refinements (Hüllermeier, 2021; Sciences et al., 2018).

These obligations converge with an independent line of work at DNV, whose probabilistic digital twin (PDT) concept formalises how epistemic and aleatoric uncertainty should be represented and propagated through a digital asset representation to support sequential, risk-informed decisions (Agrell et al., 2023). The PDT and the present framework address complementary questions rather than competing ones: the PDT supplies a rigorous calculus for carrying uncertainty from sensor through model to decision, while the typology and validation hierarchy of Sections 4 and 5 specify which safety function that decision serves and what evidence it must meet before it is trusted. A mature Type II or higher implementation would be expected to adopt PDT-style propagation as the technical substrate beneath the uncertainty-bound requirement already listed in Table 4; formal end-to-end propagation across all six layers remains an open extension of the present work, taken up in Section 7.4.

7.3. From demonstration to deployment: an organisational gap

The distance between the demonstrations of Section 6 and a trustworthy plant deployment is chiefly organisational. Brownfield facilities carry decades of instrumentation heterogeneity, fragmented documentation, and institutional memory that no data pipeline captures; the data-qualification effort of Layer 1 is, in such environments, a programme of discovery rather than verification (Di Carlo et al., 2021; Perno et al., 2022). A twin whose outputs are analytically correct but organisationally orphaned changes nothing: the barrier ranking of Demonstration B reduces risk only if a named owner receives it, a procedure converts it into inspection work orders, and an authority closes the loop. Cybersecurity adds a further coupling that process safety standards have not yet absorbed: a safety-connected twin is an attack surface through which corrupted data can be transformed into corrupted decisions, so security engineering to IEC 62443 (Cindrić et al., 2025; Leander et al., 2019) and safety engineering must proceed jointly rather than sequentially (Wang et al., 2024). Finally, no recognised qualification pathway yet exists against which a regulator or an insurer can accept a twin for safety-critical use; until standards bodies close this gap, each deployment carries the full burden of facility-specific justification, and the tiered hierarchy of Section 5.2 is offered as a pragmatic interim template (Commission, 2003; Mitov, 2025).

Human adoption deserves separate emphasis because it fails quietly. Operator trust in an advisory system is built from explainability, consistency, and a track record, and it is destroyed by unexplained alerts and false alarms far faster than it is rebuilt; the alarm-fatigue literature documents how sustained exposure to low-value notifications teaches users to ignore precisely the channel that must be heeded when it matters (Commission, 2014; EEMUA, 2013). The Layer 5 requirements of the architecture, contextual trend presentation, calibrated confidence display, and on-demand explanation in operational rather than statistical vocabulary, are therefore not interface preferences but adoption preconditions, and they demand human factors engineering and representative-condition user testing of the kind that situation-awareness research has long prescribed (Endsley, 2017). Organisational change management completes the picture: introducing a twin redistributes information, and therefore authority, among operations, maintenance, and process safety functions, and unmanaged redistribution of authority in a safety-critical setting is a hazard in its own right. Deployments that treat the twin as an IT delivery rather than as a modification to a sociotechnical safety system should expect the fate of most unmanaged modifications.

A concrete illustration of this coupling is (Liu et al., 2025) RESPONSE framework, which hardens process control against cyberattack by running a hardwired, internet-isolated offline controller in continuous parallel with the online controller that operates continuously in parallel with the online controller, rather than relying primarily on timely attack detection. The choice is instructive: even so, the framework's own case studies report a detection latency of under one hour for a simple energy-balance detector against a gradual attack on, e.g., a sensor, while a fixed five-hour window is assumed elsewhere merely to confirm and neutralise an attack once flagged, and the authors state explicitly that conventional detection cannot be assumed reliable against a sophisticated, stealthy adversary. The lesson for safety-critical digital twins is the one this paper has argued from a different direction: where a twin's protective value depends on detecting an anomaly, whether induced by process drift or cyber intrusion, within a time short enough to matter, that dependency is itself a claim requiring validation evidence (Table 4), and continuous, detection-independent redundancy, as RESPONSE illustrates, is preferable wherever it can be achieved.

7.4. Research directions

Five directions follow directly from the gaps this paper leaves open. Near term, the field needs open, fault-injected, risk-annotated benchmark datasets and reproducible reference implementations against which competing approaches can be compared, and it needs uncertainty-aware, explainable methods matured for operational time pressure. Medium term, certification frameworks that translate the hierarchy of Section 5.2 into normative requirements, and joint safety-security risk models for the cyber-physical interface, are the priorities. Long term, the frontier is escalation: twins able to project domino sequences across unit boundaries, coupling fire, blast, toxic release, and utility loss into a single fast-time projection for emergency command, a capability whose value the domino-effect literature makes clear (Gholamizadeh, Zarei, et al., 2024b; Zio, Miqueles, 2024).

Alongside these field-level directions, two improvements would advance the specific constructs demonstrated here: replacing the illustrative barrier and threat parameters with plant-calibrated values drawn from layer-of-protection analysis and reliability records, and substituting the transparent health-score functions with hybrid physics-informed and data-driven estimators fed from proof-test, inspection, and condition-monitoring histories.

7.5. Limitations

Several limitations bound the claims made here. The two demonstrations rest on synthetic data, so their results are an existence proof that the lower and middle tiers are realisable, not evidence of performance on an operating plant; transfer to field conditions remains to be established. The barrier probabilities, threat frequencies, and degradation parameters of Demonstration B are illustrative values representative of typical process-industry ranges rather than quantities calibrated to a specific facility, and the reported top-event likelihood is a relative proxy computed under assumptions of conditional independence and series combination along each threat line, so its absolute magnitude should not be read as a plant frequency. Each demonstration exercises a single unit and a single scenario, which is sufficient to expose behaviour but not to characterise robustness across process types, fault populations, or operating regimes. The digitised Bow-Tie retains static combinational logic; temporal dependency and common-cause coupling, which a Type II deployment would need, are discussed but not implemented in the demonstration. Likewise, the calibrated-uncertainty and cross-layer propagation obligations set out in Sections 5 and 7.2 are specified as requirements rather than demonstrated end-to-end. Finally, the framework has not been subjected to field validation or to any recognised qualification process, and the validation hierarchy of Section 5.2 remains a proposal rather than a standard. These limitations do not undermine the framework, but they mark the boundary between what has been shown and what must still be earned.

7.6. Industrial application scenarios

The framework applies wherever safety depends on the condition of identifiable barriers and on the timely detection of developing upsets. In refining and petrochemicals, a Type II living Bow-Tie can feed mechanical-integrity and risk-based-inspection programmes, converting continuously updated barrier-health estimates into inspection intervals and turnaround scope, while a Type I monitoring twin watches critical rotating and separation equipment. On offshore installations, where inspection access is costly and barrier management is already formalised, the same constructs support risk-informed inspection planning and continuous oversight of safety-critical elements. In liquefied-natural-gas and cryogenic-storage operations, the monitoring and risk tiers address containment integrity and overpressure or rollover precursors. Facilities central to the energy transition, including hydrogen production and storage and ammonia handling, present relatively young operating experience and sparse failure data, which is precisely the regime in which a barrier-centric twin with calibrated uncertainty is most valuable. Bulk storage terminals and tank farms, where corrosion and overfill dominate the risk, map naturally onto the threat structure of Demonstration B. For major-hazard sites more broadly, a Type IV emergency-projection twin can support incident command by projecting escalation across unit boundaries. In each setting, the type and the validation tier that the application requires follow directly from the safety function it serves, as set out in Sections 4 and 5.

8. Conclusions

This paper has presented a structured framework for digital twins whose purpose is process safety rather than production performance, together with computational evidence that its lower and middle tiers are realisable with current methods. The functional typology classifies safety twins into five types by the safety function they perform, and attaches to each type the validation evidence, autonomy boundary, and governance obligation that trust in that function requires. The six-layer reference architecture specifies the analytical chain from qualified field data to human decision and places a governance layer across the whole, while the tiered validation hierarchy transposes the graded-rigour philosophy of functional safety onto data-driven components. The two demonstrations quantify what these constructs deliver under controlled conditions: a multivariate monitoring pipeline that confirmed an abrupt fault within 8 min and slow faults within operationally useful horizons at one false alarm per day, and a living Bow-Tie that made a 21.2% erosion of protection visible and traced most of the associated risk sensitivity to a single barrier.

The framework's limitations, set out in full in Section 7.5, mark its honest boundary. The demonstrations are synthetic, and synthetic performance is an existence proof, not a deployment warranty. The validation hierarchy is a proposal, not a standard, and it will remain a facility-by-facility burden until standardisation bodies act. Above all, every capability described here is conditional on data quality, calibrated uncertainty, and governance; where those conditions fail, the twin's confidence outruns its knowledge, and confident ignorance is the one output a safety system must never produce.

For practitioners, the framework implies a definite adoption pathway rather than a leap. Begin at Type I, where the validation burden is lightest and the value of continuous multivariate surveillance is easiest to demonstrate on a single well-instrumented unit; use that deployment to build the Layer 1 data-qualification discipline and the Layer 6 governance habits that every later tier will require. Advance to Type II only when barrier records, proof-test data, and maintenance histories are current enough to feed health estimators honestly, because a living Bow-Tie built on stale records is a static Bow-Tie with a misleading interface. Treat Types III and IV as parallel tracks owned by training and emergency-preparedness functions, respectively, and treat Type V as a destination that is reached, if at all, only through functional-safety qualification of specific, narrowly bounded action pathways. At every step, let the tiered hierarchy of Table 4 define the evidence to be assembled before the next step is taken.

The enduring aims of loss prevention are unchanged: understand the hazards, keep the barriers sound, and give the people who run the plant a true picture of the risk they are managing. Digital twins, built to the discipline set out here, serve those aims by making hazard awareness continuous, barrier condition visible, and safety decisions evidence-based. Built without that discipline, they add a new and well-dressed way of being wrong. The difference lies not in the technology but in the typology, validation, and governance that surround it, and supplying that surrounding structure has been the purpose of this paper.

CRediT authorship contribution statement

Kamran Gholamizadeh: Writing – original draft, Software, Methodology, Conceptualization. **Stewart W. Behie:** Writing – review & editing. **Hans J. Pasman:** Writing – review & editing, Supervision, Conceptualization. **Ming Yang:** Writing – review & editing.

Supplementary material

The Python implementations of both demonstrations, including simulation, fault-injection, monitoring, and figure-generation code, are available in **Appendix File**.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This article is a synopsis of a chapter prepared by the authors for the fifth edition of Lees' Loss Prevention in the Process Industries (Elsevier).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psep.2026.109780.

References

Acuña, E.G.A., 2026. Artificial intelligence–driven simulation models for industrial accident prevention in chemical process engineering. Sci. J. Eng. Res. 2 (2), 234–247.

Agnusdei, G.P., Elia, V., Gnoni, M.G., 2021. Is digital twin technology supporting safety management? A bibliometric and systematic review. Appl. Sci. 11 (6), 2767.

Agrell, C., Dahl, K.R., Hafver, A., 2021. Optimal sequential decision making with probabilistic digital twins. arXiv Prepr. arXiv2103. 07405.

Agrell, C., Dahl, K.R., Hafver, A., 2023. Optimal sequential decision making with probabilistic digital twins. SN Appl. Sci. 5 (4), 114.

Al Ghazal Abdullah, H., Al Nutaifi Abdulaziz, M., Al Shaikh Ahmed, A., Altook Talal, A., 2022Saudi Aramco's Process Digital Twin: Advances & Implementation Strategy. Paper presented at the Proceedings of Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE..

Aliabadi, M.M., Gholamizadeh, K., 2021. Locating urban CNG stations using quantitative risk assessment: using the Bayesian network. Paper presented at the Safety and Reliability.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Benjamins, R., 2020. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion. 58, 82–115.

Azangoo, M., Sorsamäki, L., Sierla, S.A., Mätäsniemi, T., Rantala, M., Rainio, K., Vyatkin, V., 2022. A methodology for generating a digital twin for process industry: a case study of a fiber processing pilot plant. IEEE Access. 10, 58787–58810.

Bell, R. (2006). *Introduction to IEC 61508*. Paper presented at the Acm international conference proceeding series.

Bevilacqua, M., Bottani, E., Ciarapica, F.E., Costantino, F., Di Donato, L., Ferraro, A., Ortenzi, M., 2020. Digital twin reference model development to prevent operators' risk in process plants. Sustainability 12 (3), 1088.

CCPS, 2010. Guidelines for risk based process safety. John Wiley & Sons.

Cindrić, I., Jurčević, M., Hadjina, T., 2025. Mapping of industrial IoT to IEC 62443 standards. Sensors 25 (3), 728.

Commission, I.E. (2003). Functional safety-Safety instrumented systems for the process industry sector-Part 3, Guidance for the determination of the required safety integrity levels. *IEC 61511-3*.

Commission, I.E., 2014. IEC 62682 Management of alarm systems for the process industries. IEC, *Geneva, Switzerland*, pp. 24–26.

Correa, O.C., Mariano, J.S., Amaro, E.P., Santos, E. d N., 2023. Process safety management in oil and gas operating units through digital twin platform: a digital approach for safety control and process intervention. Paper presented at the Offshore Technology Conference Brasil.

De Dianous, V., Fievez, C., 2006. ARAMIS project: a more explicit demonstration of risk control through the use of bow–tie diagrams and the evaluation of safety barrier performance. J. Hazard. Mater. 130 (3), 220–233.

Di Carlo, F., Mazzuto, G., Bevilacqua, M., Ciarapica, F.E., Ortenzi, M., Di Donato, L., Pirozzi, M., 2021. A process plant retrofitting framework in Industry 4.0 perspective. IFAC-Pap. 54 (1), 67–72.

EEMUA, G., 2013. Alarm systems: A guide to design, management and procurement. Eng. Equip. Mater. Users Assoc.

Endsley, M.R., 2017. Toward a theory of situation awareness in dynamic systems. Situational awareness. Routledge, pp. 9–42.

Gholamizadeh, K., 2026. Prospective risk scenario intelligence for industry 5.0: a data-driven framework using synthetic futures, deep learning, and counterfactual interventions. Reliab. Eng. & Syst. Saf., 113379

Gholamizadeh, K., Pasman, H.J., 2026. HERO: a hybrid expert-relational ontology framework for robust decision-making in the risk management of socio-technical systems. Reliab. Eng. & Syst. Saf., 112825

Gholamizadeh, K., Zarei, E., BahooToroody, A., 2025. Analyzing the dynamic domino effect in fuel truck parking lots. Reliab. Eng. & Syst. Saf. 262, 111256.

Gholamizadeh, K., Zarei, E., Gualtieri, L., De Marchi, M., 2025. Advancing occupational and system safety in Industry 5.0: effective HAZID, risk analysis frameworks, and human-AI interaction management. Saf. Sci. 184, 106770.

Gholamizadeh, K., Zarei, E., Yazdi, M., Amin, M.T., 2024a. Applying Bayesian networks to safety causation analysis and modeling in socio-technical systems: bridging theory and practice. Safety causation analysis in sociotechnical systems: advanced models and techniques. Springer, pp. 363–404.

Gholamizadeh, K., Zarei, E., Yazdi, M., Amin, M.T., 2024b. Data-driven approaches for accident analysis in sociochemical systems. Saf. Causa Anal. Socio Syst. Adv. Models Tech. 457–486.

Gholamizadeh, K., Zarei, E., Yazdi, M., Ramezanifar, E., Aliabadi, M.M., 2024. A hybrid model for dynamic analysis of domino effects in chemical process industries. Reliab. Eng. & Syst. Saf. 241, 109654.

Gill, M.S., Vyas, J., Markaj, A., Gehlhoff, F., & Mercangöz, M. (2025). Leveraging LLM agents and digital twins for fault handling in process plants. Paper presented at the 2025 IEEE 30th International Conference on Emerging Technologies and Factory Automation (ETFA).

Grieves, M., Vickers, J., 2016. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. Transdisciplinary perspectives on complex systems: New findings and approaches. Springer, pp. 85–113.

Häring, I., 2021. The standard IEC 61508 and its safety life cycle. Technical Safety, Reliability and Resilience: Methods and Processes. Springer, pp. 193–207.

Hu, S., Wang, S., Su, N., Li, X., Zhang, Q., 2021. Digital twin based reference architecture for petrochemical monitoring and fault diagnosis. Oil & Gas. Sci. Technol. Rev. D. 'IFP Energ. Nouv. 76 (9).

Hüllermeier, E. (2021). Aleatoric and Epistemic Uncertainty in Machine Learning. In.

Jia, J., Wang, X., Xu, Y., Song, Z., Zhang, Z., Wu, J., Liu, Z., 2024. Digital twin technology and ergonomics for comprehensive improvement of safety in the petrochemical industry. Process. Saf. Progress. 43 (3), 507–522.

Kaarlela, T., Pieskä, S., & Pitkäaho, T. (2020). *Digital twin and virtual reality for safety training*. Paper presented at the 2020 11th IEEE international conference on cognitive infocommunications (CogInfoCom).

Keprate, A., Bagalkot, N., 2022. Use of digital twins for process safety management. In: Methods in Chemical Process Safety, 6. Elsevier, pp. 561–589.

Khan, F., Rathnayaka, S., Ahmed, S., 2015. Methods and models in process safety and risk management: Past, present and future. Process. Saf. Environ. Prot. 98, 116–147.

Kletz, T., Amyotte, P., 2019. What went wrong?: case histories of process plant disasters and how they could have been avoided. Butterworth-Heinemann.

Leander, B., Čaušević, A., & Hansson, H. (2019). *Applicability of the IEC 62443 standard in Industry 4.0/IIoT*. Paper presented at the Proceedings of the 14th International Conference on Availability, Reliability and Security.

Lee, J., Xu, C., Xie, Y., 2025. Flow-based Conformal Prediction for Multi-dimensional Time Series. arXiv Prepr. arXiv2502. 05709.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J., 2018. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech. Syst. Signal. Process. 104, 799–834.

Li, X., Zhang, Y., Li, C., Wang, T., Xi, C., 2022. Application of digital twin in handling and transportation of hazardous chemicals. Appl. Sci. 12 (24), 12746.

Liu, L., Sajid, Z., Kravaris, C., Khan, F., 2025. RESPONSE-a resilient framework to manage cyber-attacks on cyber-physical process systems. Digit. Chem. Eng. 15, 100238.

Mane, S., Dhote, R.R., Sinha, A., Thirumalaiswamy, R., 2024. Digital twin in the chemical industry: A review. Digit. Twins Appl. 1 (2), 118–130.

Mitov, N., 2025. Digital Twins in Functional Safety: Classification, Applicability, Tool Qualification, Model Architecture, and Validation Methodology. Paper presented at the 2025 XXXIV International Scientific Conference Electronics (ET).

Mohammadfam, I., Kalatpour, O., Gholamizadeh, K., 2020. Quantitative assessment of safety and health risks in HAZMAT road transport using a hybrid approach: a case study in Tehran. ACS Chem. Health & Saf. 27 (4), 240–250.

Mohammadfam, I., Zarei, E., Yazdi, M., Gholamizadeh, K., 2022. Quantitative risk analysis on rail transportation of hazardous materials. Math. Probl. Eng. 2022 (1), 6162829.

Mohd Sahak, M.Z., Myo Thant, M.M., Tumian, A., Harun, Z., Castillano, E., Chee, S.B., Tan, K., 2022. A Holistic Approach to the Development of Digital Twin Advisor for Upstream Production Process. Paper presented at the Abu Dhabi International Petroleum Exhibition and Conference.

Muhlheim, M.D., Ramuhalli, P., Guler Yigitoglu, A., Huning, A., Wood, R., Narvaez, J.L., Saxena, A., 2025. First-of-a-Kind Risk-Informed Digital Twin for Operational Decision Making. Nucl. Sci. Eng. 199 (11), 1899–1914.

Murray, P., Wattis, Z., Bain, B., Golowczynski, M., Sadd, A., 2019. Towards a digital twin supporting risk based decision making for offshore installations. Paper presented at the SPE Offshore Europe Conference and Exhibition.

Pal, P.K., Hens, A., Behera, N., Lahiri, S.K., 2025. Digital twins: Transforming the chemical process industry—A review. Can. J. Chem. Eng. 103 (8), 3611–3636.

Paltrinieri, N., Comfort, L., Reniers, G., 2019. Learning about risk: Machine learning for risk assessment. Saf. Sci. 118, 475–486.

Paltrinieri, N., Khan, F.I., 2016. Dynamic risk analysis in the chemical and petroleum industry: evolution and interaction with parallel disciplines in the perspective of industrial application. Butterworth-Heinemann.

Park, J.-S., Lee, D.-G., Jimenez, J.A., Lee, S.-J., Kim, J.-W., 2023. Human-focused digital twin applications for occupational safety and health in workplaces: a brief survey and research directions. Appl. Sci. 13 (7), 4598.

Pasman, H.J., 2015. Risk analysis and control for industrial processes-gas, oil and chemicals: a system perspective for assessing and avoiding low-probability, high-consequence events. Butterworth-Heinemann.

Perno, M., Hvam, L., & Haug, A. (2020). *Enablers and barriers to the implementation of digital twins in the process industry: A systematic literature review*. Paper presented at the 2020 IEEE international conference on industrial engineering and engineering management (IEEM).

Perno, M., Hvam, L., Haug, A., 2022. Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers. Comput. Ind. 134, 103558.

Peterson, L., Gosea, I.V., Benner, P., Sundmacher, K., 2025. Digital twins in process engineering: An overview on computational and numerical methods. Comput. & Chem. Eng. 193, 108917.

Qamsane, Y., Phillips, J.R., Savaglio, C., Warner, D., James, S.C., Barton, K., 2022. Open process automation-and digital twin-based performance monitoring of a process manufacturing system. IEEE Access. 10, 60823–60835.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.

Rasheed, A., San, O., Kvamsdal, T., 2020. Digital twin: Values, challenges and enablers from a modeling perspective. IEEE Access. 8, 21980–22012.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Mayo, D., 2020. Five ways to ensure that models serve society: a manifesto. Nature 582 (7813), 482–484.

Sciences, N.A. o, Medicine, Engineering, D. o, Sciences, P., Aeronautics, Board, S.E., Committee, A.S.A., 2018. In-time aviation safety management: Challenges and research for an evolving aviation system. National Academies Press.

Singh, P., van Gulijk, C., Sunderland, N., 2024. The BowTie as a digital twin: how a BowTie looks different from a data perspective. Safety 10 (2), 34.

Sklet, S., 2006. Safety barriers: Definition, classification, and performance. J. Loss Prev. Process. Ind. 19 (5), 494–506.

Smith, D.J., Simpson, K.G., 2020. The Safety Critical Systems Handbook: A Straightforward Guide to Functional Safety: IEC 61508 (2010 Edition), IEC 61511 (2016 Edition) Also Related Guidance on Cyber Security & Including Machinery and Other Industrial Sectors. Elsevier.

Tao, F., Zhang, H., Liu, A., Nee, A.Y., 2018. Digital twin in industry: State-of-the-art. IEEE Trans. Ind. Inform. 15 (4), 2405–2415.

Vessel, H.H.-P. (2020). *A digital twin for safety and risk*. Paper presented at the Designing for Digital Transformation. Co-Creating Services with Citizens and Industry: 15th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2020, Kristiansand, Norway, December 2–4, 2020, Proceedings.

Wang, R., Venugopalan, S., & Adepu, S. (2024). *Safety analysis for cyber-physical systems under cyber attacks using digital twin*. Paper presented at the 2024 IEEE International Conference on Cyber Security and Resilience (CSR).

Wu, S., Gong, H., Yu, L., Zhang, A., Zhang, L., Liu, Y., 2025. Physics-informed dynamic Bayesian networks for time-dependent reliability prediction of subsea wellhead sealing system with multi-states. Eng. Appl. Artif. Intell. 162, 112492.

Wu, S., Hu, Y., Zhang, L., Liu, S., Xie, R., Yin, Z., 2024. Intelligent risk identification for drilling lost circulation incidents using data-driven machine learning. Reliab. Eng. & Syst. Saf. 252, 110407.

Zabarin, I., & Shilin, D. (2023). *Development of the digital twin for a multi-purpose processing plant training simulator*. Paper presented at the 2023 5th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE).

Zarei, E., Gholamizadeh, K., Khan, F., Khakzad, N., 2022. A dynamic domino effect risk analysis model for rail transport of hazardous material. J. Loss Prev. Process. Ind. 74, 104666.

Zio, E., Miqueles, L., 2024. Digital twins in safety analysis, risk assessment and emergency management. Reliab. Eng. & Syst. Saf. 246, 110040.