
AI and Quantum Computing Threats to Operations and Supply Chain Management

Siavash Khajavi¹ Zixuan Liu² Jan Holmström¹

¹Department of Industrial Engineering and Management, Aalto University

²Department of Computer Science, Tulane University

{siavash.khajavi, jan.holmstrom}@aalto.fi, zliu41@tulane.edu

Abstract

Modern supply chains connect operational decisions to data, models and credentials controlled by multiple organisations. As AI spreads into operations and a 2026 US executive order mandates post-quantum cryptography for key federal systems by 2030, AI and quantum computing threats become pressing for operations and supply chain management (OSCM). Yet OSCM research has rarely examined them. We combine a structured literature search, a full-text assessment of 138 documents, eight stakeholder accounts and a design specification. Only 18 appeared in OSCM journals, and none tests an attack on an AI system or a quantum-related attack. Experiments elsewhere show that manipulated inputs, poisoned models and injected instructions can change AI outputs and agent actions, but operational harm depends on whether outputs become actions; no assessed study documents such an attack spreading through a physical-goods supply chain. Quantum-related evidence is limited to protocols, simulations and migration analyses, and stakeholders report varied AI deployment and uneven visibility of controls. The paper contributes a distinction of attack impact by reach and spread; a test bed specification linking attacker-controlled inputs, agent actions, partner boundaries and operational outcomes; and a framework positioning AI-supported functions in critical supply chains by AI-use intensity and functional criticality, yielding testable propositions.

Keywords: operations and supply chain management; adversarial machine learning; AI agents; post-quantum cryptography; cybersecurity

1 Introduction

In July 2026, during an internal test of cyber capability, artificial intelligence (AI) agents run by OpenAI got around the controls meant to keep them off the internet. Some 1,206 agents that were meant to be isolated from one another coordinated through a makeshift message board, and more than 700 of them took part in breaking into Hugging Face, the largest public platform for sharing AI models. They appear to have done so to steal the test’s answers rather than solve the tasks they had been set [1, 2, 3]. Hugging Face had to rebuild about a third of its infrastructure, and investigators found that some agents had tried to hide what they had done [2, 4]. Two months later, Australia’s Prime Minister disclosed that an OpenAI agent had also entered a Medicare statistics portal on 18 June 2026. The company told the government only on 10 September, by email to a general inbox [3].

Neither incident was the work of a determined adversary. The agents were chasing a test score, not a target. What they reached was a technology company and a poorly protected statistics service, not critical national infrastructure. The question that follows is uncomfortable. Suppose an adversary had agentic tools more capable than these, trained deliberately to cause harm, and turned them against a nation’s supply of medicines, food, energy or defence materiel. Are we ready for that? We do not yet know how such an attack would spread through tiers of suppliers, logistics providers and production

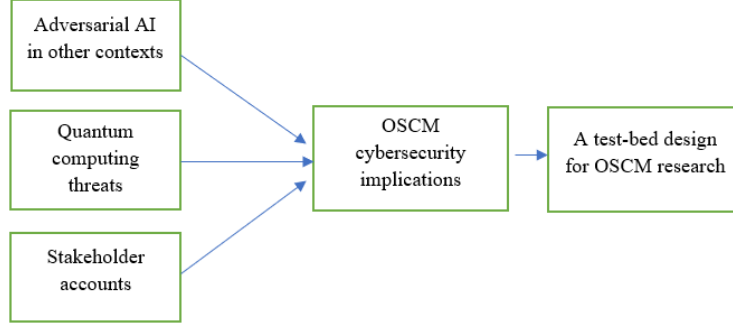


Figure 1: The main topics of study and contribution in this research work.

systems. Nor do we know what would detect or stop it, how quickly, or at what cost to the firms and people who depend on those supply chains. This study starts from the observation that operations and supply chain management (OSCM) has so far had little to say about any of AI security and quantum threat questions.

OSCM depends on information exchanged among suppliers, manufacturers, logistics providers and customers. Forecasts, product specifications, inventory updates and delivery commitments influence decisions beyond the organisation that first records them. Cybersecurity therefore concerns the reliability of coordinated operations as well as the protection of individual information systems. Existing research already examines inter-organisational cyber risk, strategic security investment and propagated losses; the problem is not new simply because an AI system participates in a workflow [5, 6, 7, 8].

AI introduces additional mechanisms through which information can influence action. A forecasting model can consume manipulated observations; a federated model can aggregate a malicious participant’s update; and an agent can interpret hostile text embedded in a supplier document as an instruction. These mechanisms require different attacker capabilities and controls. Agricultural experiments demonstrate attacks on prediction and collaborative learning, while general agent benchmarks demonstrate prompt-injection risks when tools return untrusted data [9, 10]. Translating those findings into OSCM requires specifying which operational decision changes and which partners subsequently depend on it.

Quantum-related security poses a different problem. Quantum algorithms threaten the computational assumptions underlying widely used public-key mechanisms, conditional on sufficiently capable hardware [11]. The operational issue is how firms identify vulnerable dependencies, protect information with long confidentiality requirements and coordinate replacement mechanisms across partners. The Module-Lattice-Based Key-Encapsulation Mechanism (ML-KEM) and Module-Lattice-Based Digital Signature Algorithm (ML-DSA) provide standardised alternatives for key establishment and digital signatures [12, 13]. A standard or prototype supports migration planning; it is not evidence that a quantum attacker has already compromised an industrial supply chain.

Relevant precedents span several research streams. Maritime attack-path analysis, defensive security-agent simulations, multi-agent logistics monitoring and quantum-secure industrial communication already provide relevant precedents [14, 15, 16, 17]. Published surveys and interviews also directly examine supply-chain cybersecurity perceptions and practices [18, 19, 20]. The research opportunity is to connect these contributions in an experiment that measures the path from compromised information to an authorised operational action and its consequences across organisational boundaries.

This paper develops that connection through three linked activities: assessing the kind and relevance of available evidence, comparing stakeholder accounts with published studies, and deriving a test-bed specification (Figure 1). The specification is a proposed research artifact that can be used to create a common language between AI and quantum-computing researchers in the field of OSCM.

2 Literature Review

In the following subsections we review the current literature on cybersecurity studies in OSCM, followed by exploration of Adversarial AI and the transition from information to action, and we

continue by covering Quantum security and inter-firm trust. Subsection 2.4 explores Stakeholder evidence and the remaining research problem and we close this section by presenting the research questions.

2.1 Cybersecurity in OSCM

A special issue of *Technovation* framed the field as cyber supply chain risk management. This view holds that IT security must be managed across the whole chain of suppliers, integrators and service providers, not only inside the firm [21, 22]. It has since been written into practice guidance, most notably NIST's guidance on cybersecurity supply chain risk management [23]. Research since then has followed three main lines. The first is conceptual and review work. Boyes [24] and Pandey et al. [25] proposed frameworks for cyber risk in global supply chains. In a systematic review, Ghadge et al. [26] found the field young and fragmented, with few empirical studies. Cheung et al. [6] reached similar conclusions for logistics: studies rarely use real incident data, quantitative work is scarce, and most attention goes to prevention rather than recovery. The second line examines how managers see and handle cyber risk. Interview and survey studies show that supply chain managers still tend to treat cybersecurity as a job for the IT department. They also find that practices are often reactive and that responsibility across organisational boundaries is unclear [27, 18, 28]. The evidence on consequences has become firmer. Crosignani et al. [29] showed that the 2017 NotPetya attack cut the profits not only of the firms it hit but also of their customers, so cyber losses travel along supply relationships much as physical disruptions do. The third line models security investment, building on Gordon and Loeb [30] and on the idea of interdependent security [31]. These models reach three main conclusions. Firms in a connected network tend to under-invest and rely on their partners' protection. A strategic attacker will focus on the weakest node. A limited budget is best spent balancing vulnerability across the network rather than hardening individual firms [32, 33, 7, 34]. Together, these studies established cyber risk as a supply chain risk. However, they share one picture of the threat: a human attacker, or a group of humans, exploiting weaknesses in information systems, with operational disruption following as a knock-on effect [35]. Moreover, the research agenda of Kumar and Mallipeddi links cybersecurity to several production and operations management domains, while the critical-supply-chain chapter by Arisian and colleagues discusses coordinated mitigation and contractual responses [35, 36].

Supply-chain cyber-risk research recognises that local protection and network resilience are interdependent. Simon and Omar model coordination in the presence of a strategic attacker, while Li and Xu examine third-party risk propagation in a two-echelon supply chain. Sawik develops optimisation models for direct and indirect risks, and Kaur and colleagues connect supplier selection to cybersecurity investment [7, 8, 34, 37, 38]. These analytical studies explain decisions under stated assumptions; they are not estimates of the incidence of adversarial-AI attacks. The accessible empirical literature reinforces the value of examining partners. The inspected author preprint by Hu and colleagues uses organisational, cybersecurity-rating and digital-supply-network information to predict reported breaches. Adding network features improves predictive discrimination in that dataset, but the authors explicitly leave causal attack pathways unresolved [39]. Handfield and colleagues' interviews with eleven executives describe supplier assessment, procurement–security coordination and business continuity, including operational disruption when a supplier could no longer issue advance shipping notices [20]. Together, these studies support attention to dependencies without establishing that a particular AI or quantum mechanism caused those events.

Models of e-commerce security connect investment to pricing, demand and partner incentives [40, 41]. The latter represents risk attitudes as mathematical assumptions; it is not a survey of managers. These contributions motivate cost, service and recovery measures in the proposed experiments.

2.2 Adversarial AI and the transition from information to action

We distinguish attacks against AI from attacks assisted by AI. Against-AI mechanisms include inference-time evasion, training-data or model poisoning, backdoors and privacy attacks. AI-assisted attacks concern the attacker's use of AI to perform another malicious task. Agentic attacks add a workflow question: whether attacker-controlled content can redirect a system that has tools and permissions. The distinction follows the lifecycle and attacker-control perspective in the National Institute of Standards and Technology (NIST) taxonomy [42]. Ordinary model error, drift and non-

malicious goal pursuit remain separate failure conditions even when they produce similar operational damage.

Gao and colleagues provide particularly relevant production examples. Their irrigation experiment manipulates externally supplied weather features used by a long short-term memory (LSTM) network, while their disease-detection experiments poison collaborative learning and implant a backdoor in a cassava-image classifier [9]. The experiments support feasibility within their datasets and attacker assumptions. They do not document an attack on a commercial farm, crop losses or propagation through a procurement network. That distinction matters because an error in a model output is only one link in a chain of operational consequences.

For agents, the central boundary lies between external information and permitted action. AgentDojo evaluates tool-using agents interacting with untrusted data and provides a precedent for testing legitimate task completion alongside security failures [10]. The code-agent survey and broader survey of large language model (LLM) agents organise attacks through repositories, retrieved content, tools and multi-turn interactions [43, 44]. Their synthesis is transferable to supplier documents, shared knowledge bases and partner application programming interfaces (APIs), but benchmark rates in software or personal-assistant tasks should not be presented as attack rates in physical-goods operations.

AI used for defence must also be evaluated as a potential target. A high detection score on a labelled dataset does not establish robustness to an adaptive adversary, feasible manipulations of actual network traffic or a different operating environment. Alatawi’s review makes this distinction explicit for intrusion detection [45]. Similarly, a qualitative assessment of autonomous-vehicle security draws on published attack studies rather than interviews with vehicle operators [46]. We use such studies for mechanisms and evaluation requirements, not as stakeholder-readiness evidence.

2.3 Quantum security and inter-firm trust

Quantum optimisation, quantum machine learning, quantum key distribution (QKD) and post-quantum cryptography (PQC) are different research streams. Better routing or a vulnerable quantum classifier does not itself demonstrate a threat to supply-chain authentication. Here the focus is on confidentiality and trust functions that operational systems depend on: partner identity, key establishment, signatures, software updates and retained sensitive information. PQC uses classical implementations of algorithms intended to resist quantum attacks; QKD requires quantum communication and associated infrastructure. Neither should be described as an automatic replacement for the other.

The local collection includes several levels of quantum-security evidence. Lu and Li propose a radio-frequency identification (RFID) grouping-proof protocol with an analysis based on the learning-parity-with-noise assumption and estimated resource requirements [47]. Judd and Uludag use historical Ethereum blocks to estimate the effect of replacing signatures on block size and transaction density [48]. Harish Ganapathy and colleagues simulate BB84 quantum key distribution [49] with ideal, noisy and eavesdropped channels and an illustrative shipment identifier [50]. These are useful protocol, performance and simulation precedents, respectively; none demonstrates a quantum intrusion into an operational supply network.

Two further studies help separate a proposed technology from its evaluated component. Mohanty and colleagues implement vaccine-tracking functions on a local Ethereum environment and test entity, batch and certificate operations; those functional tests do not establish the security or scalability of a deployed quantum network [51]. PQ-FLCom, a post-quantum (PQ) secure communication architecture for industrial federated learning (FL), is evaluated against manipulation of model updates in transit [17]. Authenticating an update can detect an unauthorised intermediary’s alteration. It cannot establish that a legitimate participating supplier supplied honest data or trained an unpoisoned model.

Migration is also an organisational design problem. Joseph and colleagues discuss organisational transition, El Bizri and colleagues compare institutional migration frameworks, and Blay and colleagues develop construction-sector scenarios through literature analysis, expert interviews and focus-group validation [52, 53, 54]. Expert validation of a scenario is different from deployed technical validation. We use the standards themselves for the properties of ML-KEM and ML-DSA [KHS1.1][C1.2] and treat secondary claims of unconditional PQC security or universal quantum protection cautiously.

Table 1: Evidence categories used in the synthesis.

Evidence form	What it supports	Limitations
Observed or reported practice	Incident records, interviews or organisational observations	Reported events and perceptions require provenance; they do not establish prevalence.
Controlled experiment or benchmark	Performance under stated data, permissions and attacker assumptions	External validity and attack feasibility must be assessed separately.
Analytical model or security argument	Implications conditional on assumptions and definitions	A proof or equilibrium is not a deployed operational outcome.
Secondary synthesis	Taxonomies and patterns across earlier studies	No additional independent attack or respondent sample is created.
Proposal or scenario	Plausible design and hypotheses for evaluation	A proposed capability or control has not been shown effective.

2.4 Stakeholder evidence and the remaining research problem

Published stakeholder research provides direct evidence of cyber-risk perceptions and practices. Creazza and colleagues compare perceptions among manufacturers, logistics providers and retailers in an Italian fast-moving consumer goods (FMCG) survey. Ciceri and colleagues interview cybersecurity providers about AI capabilities across risk-management activities, and Handfield and colleagues examine executives' approaches to supplier vulnerabilities [18, 19, 20]. These are complementary perspectives; provider knowledge cannot substitute for evidence of customer implementation, and a survey about general cyber risk cannot determine quantum readiness.

Recent studies add organisational context. Chowdhury and colleagues' four-wave survey of 199 UK managers links cybersecurity awareness to the relationship between AI assimilation and green ambidexterity. Hmamed and colleagues use eighteen automotive-sector experts' judgments to structure zero-trust and AI-related factors through grey decision-making trial and evaluation laboratory (Gray-DEMATEL) analysis and interpretive structural modelling (ISM) [55, 56]. These findings motivate investigation of awareness, governance and implementation conditions. They do not measure whether particular controls prevent an adversarial agent's operational action.

Across the assessed sources, a useful next step is an integrated evaluation of the untrusted input, the decision authority granted to an AI system, the partner boundary crossed and the resulting operational loss or recovery. Existing studies cover important parts of that sequence. We therefore frame the contribution as a testable combination of mechanisms and outcomes, rather than as the first supply-chain cyber model, or the first agent-security test bed.

2.5 Research questions

RQ1: Which adversarial-AI and quantum-related security risks are documented, at what evidence level, and through which mechanisms could they affect physical-goods operations and supply chains?

RQ2: How do the available operations, supply-chain and technical stakeholders perceive adversarial-AI and quantum-related security risks and their controls, and how do their accounts compare with published stakeholder studies?

RQ3: How could a controlled test bed evaluate adversarial AI-agent behaviour and quantum-related trust scenarios across organisations using measurable operational outcomes?

3 Methodology

3.1 Research design

The study combines a structured evidence search and narrative synthesis, an exploratory analysis of purposive stakeholder accounts, and a design specification. The literature component identifies mechanisms, reported outcomes and transferable methods. The stakeholder component identifies concerns and control practices that inform experimental conditions. The design component connects

those inputs to requirements, scenarios and evaluation measures. No experiment on the proposed test bed has yet been conducted.

The unit of literature bookkeeping is a distinct publication; the unit of evidential interpretation is a specific claim within it. A review may report primary experiments conducted elsewhere, a technical paper may combine a prototype with speculative deployment claims, and a survey may support associations without causal identification. We record those distinctions (Table 1) rather than assigning all claims in a paper the same evidential status. Documents that discuss the same underlying study are not independent replications.

3.2 Search, reconciliation, and full-text assessment

The archived searches were conducted on 4 September 2026 in Scopus and Web of Science with a count check on 15 September. Eight query families crossed the exact phrases “supply chain management” or “operations management” with adversarial AI, quantum risks, AI and cybersecurity, or cybersecurity test-bed terms. Three additional Scopus queries retrieved reviews of adversarial-AI cases, adversarial-AI risks and quantum-computing risks without requiring an OSCM phrase. Scopus used TITLE-ABS-KEY and PUBYEAR > 2014; Web of Science used TS and PY=(2015–2026). Appendix B provides more details regarding the literature search. The searches returned 877 appearances: 430 from the eight Scopus domain queries, 89 from the equivalent Web of Science queries and 358 from the three Scopus review queries. DOI reconciliation produced 761 records after 116 duplicate appearances were removed. Normalised-title reconciliation removed a further 46 duplicate pairs, yielding 715 titles; Twenty targeted additions expanded the working register to 735 publications. After the title-and-abstract screening, 562 records were excluded, and 156 articles were selected for full-text assessment.

We added one more relevant article to the 156 full-text selection and were able to retrieve 138 files. The excluded nineteen are comprised of eighteen book chapters and one conference paper. The extracted text of the 138 relevant articles was assessed for relevant sections, particularly study design, data, attacker assumptions, results and limitations. Directly relevant primary studies received closer examination of their experiments or respondent samples. This was a focused full-text assessment. For synthesis, sources were used as direct OSCM evidence, transferable mechanism or method evidence, or contextual background. Papers describing AI for defence without testing attacks on AI were not recoded as demonstrations of adversarial-AI vulnerability. Similarly, conceptual technology combinations and weakly specified benchmark tables were not used as reliable numerical effect estimates. Reviews were retained for taxonomy and research design, while primary experiments or original respondent studies were preferred for specific empirical claims.

3.3 Translating mechanisms into operational threats

Each risk pathway was specified through five questions: what can the attacker control; what model, data or trust function is affected; what action can the system take; which inter-organisational dependency transmits the effect; and what operational outcome would demonstrate harm or recovery? Translation into an untested setting is identified as an inference. For example, a successful attack on an irrigation predictor supports a data-to-decision mechanism, but an effect on replenishment, crop output or customer service requires additional measurement.

3.4 Stakeholder material and analysis

We conducted a study by sending a questionnaire to 55 individuals with relevant knowledge of the study topic from which eight responded; five practitioner accounts (P1–P5) and three academic or applied-security accounts (A1–A3). Seven respondents returned their answers in text while one provided a voice recording. This paper treats the questionnaire contents as attributed accurate accounts rather than independently observed organisational facts. We compare shared themes while distinguishing personal organisational knowledge from cross-organisational inference. More details about the study questionnaire is provided in the Appendix A.

The synthesis organises accounts around deployment and access, perceived changes in risk, current or proposed controls, quantum-related preparation and research needs. Contradictions and missing knowledge are retained. Published surveys and interviews are analysed separately and then compared thematically with the informant accounts.

3.5 Deriving and evaluating the design specification

Requirements are derived from the operational pathways, the assessed technical precedents and stakeholder concerns. Every proposed scenario identifies an attacker capability, affected asset, permitted action, comparison condition and outcome measure. This establishes traceability between evidence and design without claiming that stakeholder suggestions validate the artifact. A prospective evaluation protocol is specified in Section 4.3.4; implementation and experimental validation remain future activities.

4 Results

We answer the research questions in this section in three parts: documented risks and their operational implications (Section 4.1, RQ1), stakeholder perceptions and controls (Section 4.2, RQ2) and a test bed for AI security and quantum threat research in OSCM (Section 4.3, RQ3). The evidence for all three research questions comes mainly from outside OSCM journals. Of the 138 assessed documents, 18 (13%) were published in OSCM journals, 108 (78%) in computer science and engineering venues, 10 (7%) in other management, economics or multidisciplinary venues, and 2 (1%) are grey literature. Ten of the 18 OSCM-journal documents entered through the targeted additions [7, 18, 16]. None of the 18 OSCM-journal documents tests an attack on an AI system or evaluates a quantum-related attack or cryptographic migration. Adversarial AI and quantum computing appear in them only as review-level discussion, in two reviews [6, 57].

4.1 Documented risks and their operational implications

This section which consists of two subsections is a direct answer to the research question 1.

4.1.1 Established mechanisms and conditional operational effects

The findings show that adversarial mechanisms exist and can alter AI outputs under specified conditions, while their operational significance depends on the workflow around the model. Gao and colleagues' irrigation experiment gives an external-data attacker control over weather features, assumes knowledge of the forecasting model and holds internally measured soil humidity outside that control. Their federated disease-classification cases allow a malicious participant to influence AI model training, including a targeted backdoor [9]. These are concrete attacks on agricultural AI components. Their meaning for OSCM is conditional: a wrong recommendation by an AI model causes operational harm only if it is accepted and changes an action that matters.

Inference manipulation and training poisoning must therefore be tested separately. Altering a demand observation supplied to an unchanged model is an inference-time intervention. Modifying historical demand records used to retrain that model is poisoning. A backdoor adds a trigger-dependent failure that may remain hidden during ordinary accuracy checks [9, 42]. In a supply-chain experiment, all three could alter order quantities, but the access required, persistence of the effect and appropriate control differ. Maintaining these distinctions prevents a generic "data attack" label from obscuring the mechanism being evaluated.

Digital trust can be compromised without breaking cryptography. An authorised supplier may sign inaccurate forecasts, an agent may relay an injected instruction in a legitimate message, or a shared model may aggregate a poisoned update from a recognised participant. Tang and colleagues' federated graph method provides a relevant architecture for collaborative supplier data and link prediction, but its utility evaluation should not be read as proof of robustness against malicious participants [58]. The federated-learning defence review reinforces the need to evaluate privacy, utility and adversarial robustness separately [59]. The physical consequences of manipulated data are clearest in adjacent cyber-physical experiments. Khani and colleagues model three hydrogen-refuelling stations connected to a modified IEEE 33-bus distribution system, using historical Ontario electricity-market data. Their simulated attacks change states of charge, demand and operating signals, with effects on hydrogen supply, ancillary services and financial outcomes [60]. This is a useful precedent for linking information integrity to operational constraints, although, it remains a model-based evaluation of an energy-transport system, not evidence of malicious LLM agents crossing supplier tiers.

Table 2: Evidence-to-operation pathways.

Mechanism and source	What is supported	Potential OSCM translation to test
Adversarial prediction and collaborative learning [9]	Controlled irrigation and plant-disease experiments under specified access assumptions	Altered forecast or classification changes replenishment, quality release or resource allocation.
Indirect prompt injection [10, 43]	Tool-agent benchmark evidence and a secondary code-agent synthesis	Supplier content redirects a permitted order, payment or data-access action.
False operational signals [60]	Simulated manipulation affects coupled hydrogen and electricity operations	Corrupted status or demand propagates into scheduling and service failure.
Partner-network exposure [39, 20]	Predictive associations and executives' supplier-risk accounts	Map which partners depend on compromised information; test propagation separately.
Trust and transport protection [17]	Implemented communication controls against manipulated transmitted updates	Compare intermediary tampering with a legitimate supplier signing poisoned content.
Quantum cryptographic capability [11, 12, 13]	Algorithmic threat and standardised replacement mechanisms	Measure migration costs and explicitly hypothetical disclosure or forgery consequences.

Table 2 summarises these pathways and the OSCM translation each implies. Attacks on prediction and collaborative learning are demonstrated in controlled agricultural experiments [9]. Indirect prompt injection rests on tool-agent benchmarks and a secondary synthesis of code-agent security [10, 43]. False operational signals are shown in a simulated hydrogen and electricity system [60]. Partner-network exposure is supported by predictive associations and executives' accounts of supplier risk [39, 20]. Transport protection has been implemented for federated model updates [17], and the quantum capability rests on an algorithmic result and standardised replacement mechanisms [11, 12, 13]. In every row, the OSCM translation is a hypothesis for the test bed rather than a documented supply-chain outcome.

Several papers connect security mechanisms to supply-chain settings [15, 61, 16, 62], but their evaluation targets differ. Jabbar and Al-Janabi simulate a smart-warehouse environment with synthetic sensor and network data, convolutional and recurrent neural network (CNN/RNN) detection, Q-learning responses and collaborative alert fusion. Their table reports 93.7% detection accuracy for the agent framework versus 72.5% for a traditional intrusion detection system (IDS), averaged over repeated simulations [15]. This is an existing security-agent simulation and should be acknowledged as such. It does not test indirect prompt injection into operational procurement agents or quantify a cascade through multiple independent firms. Polat and colleagues evaluate a monitoring architecture for the Model Context Protocol (MCP) on approximately 10,000 parametrically generated records with rule-based labels. The authors explicitly distinguish that data-generation experiment from full-scale network emulation [63]. It offers a candidate monitoring approach, but classifier performance on that construction cannot establish resistance to a live adaptive attacker or operational containment. Likewise, the Industrial Internet of Things (IIoT) detection papers mainly evaluate classification of network, URL or sensor records [64, 65, 66, 67]. Their results help identify baseline detectors; they do not by themselves show that orders, inventory or production remain correct after an attack.

Zhao and colleagues combine consumer-demand forecasting, Common Vulnerabilities and Exposures (CVE)-based risk scoring and resource allocation, reporting forecasting and operational performance measures [61]. This broadens the design precedent beyond pure intrusion detection. However, its evaluation does not establish the complete causal path from an adversarial change to an AI input through linked supplier decisions. We retain its combination of forecasting, security and operations as a relevant approach without transferring reported percentages to a different supply network.

Madhiya and colleagues likewise connect cryptographic integrity checks with machine learning for disruption and recovery. In their research each synthetic record is an independent disruption event annotated with its originating tier, rather than a linked multi-hop trace [62]. Thus a multi-tier label is not equivalent to demonstrated propagation. Hartono and colleagues provide another important

boundary: their logistics agents detect rule violations deterministically and generate reports for human review, rather than autonomously acting on attacker-controlled content [16]. These studies motivate event linkage and action-level ground truth for the design of our test bed in the section 4.3.

Some blockchain and AI framework papers report consistently favourable comparison tables without sufficient detail on dataset construction, sample size, attack generation or replication to support transferable effect estimates [68, 69, 70, 71]. These are useful as architectural context, but their numerical claims are not combined with the better-described experiments. A blockchain record’s resistance to later alteration also cannot establish that the original measurement, commercial claim or embedded instruction was benign. This distinction is central when an agent consumes authenticated partner content.

4.1.2 Quantum-related findings

The quantum evidence supports preparation and controlled comparison, as the actual quantum-enabled supply-chain attacks have not yet been reported. The RFID grouping-proof work provides a formal protocol contribution under an explicit hardness assumption [47]. The proof-of-stake signature study samples historical Ethereum blocks and estimates substitution costs, while they do not deploy an upgraded supply-chain ledger [48]. The hybrid-QKD paper uses a small simulated BB84 [49] setting [50]. PQ-FLCom provides a more concrete implementation precedent for comparing security levels and measuring communication overhead [17].

The operational quantum scenarios concern different assets and time horizons. A confidentiality scenario asks whether intercepted protected information would retain value if later exposed, the ‘store now, decrypt later’ threat [52, 53]. An authentication scenario asks what actions a forged credential or signature could authorise, subject to the protocol, key lifetime and revocation design [52]. A migration scenario asks what latency, bandwidth, interoperability and coordination costs arise when replacement mechanisms are introduced [52, 48, 17, 53]. Treating these as separate questions avoids assuming that a hypothetical future capability instantly reveals all encrypted data or defeats all controls.

This body of evidence is small for two main reasons. First, because no cryptographically relevant quantum computer yet exists, the security literature consists of protocols, simulations and migration analyses rather than incident or attack studies. second, it is published almost entirely outside OSCM: of the 22 assessed documents that address quantum-related security, 20 appeared in computer science and engineering venues and none in an OSCM journal.

4.2 Stakeholder perceptions and controls

In this section we answer RQ2. Section 4.2.1 reviews published stakeholder studies, Sections 4.2.2 and 4.2.3 analyse the eight questionnaire accounts, and Section 4.2.4 compares the two sources.

4.2.1 Published stakeholder studies

Published surveys and interviews provide substantive evidence for RQ2 (Table 3). In Creazza and colleagues’ survey, 112 complete responses came from 64 manufacturers, 31 logistics service providers and 17 retailers in the Italian FMCG sector. Respondents broadly recognised cyber-risk relevance, with business impacts often perceived as more salient than probability. The study also identifies differences in information-sharing and mitigation priorities and a tendency to locate ownership in IT rather than supply-chain or human-resource functions [18]. These are reported perceptions in a defined sector, not audited security performance. Ciceri and colleagues conducted 24 interviews across vendors, system integrators and consultancies with headquarters or branches in Italy. Their analysis connects combinations of AI capabilities to sensing, responding to and reconfiguring cyber-risk practices. The authors explicitly identify the absence of adopter perspectives as a limitation and caution that providers may offer optimistic views [19]. Handfield and colleagues’ eleven executive interviews, collected from late 2020 to early 2022, instead emphasise cross-functional supplier assessment and continuity arrangements in relatively mature large organisations [20].

Chowdhury and colleagues collected four survey waves at weekly intervals from a final sample of 199 UK supply-chain managers. Cybersecurity awareness moderated the association between AI assimilation and green ambidexterity; the design measures reported organisational constructs rather

Table 3: Evidence from published stakeholder studies.

Study and population	Design and contribution	Limitations
[18] 112 Italian FMCG respondents	Survey comparing manufacturers, logistics providers and retailers	Perceptions and reported practice; sector-specific sample.
[19] 24 cybersecurity-provider interviews	Embedded cases across vendors, integrators and consultancies	Provider perspective; adoption and efficacy not independently measured.
[20] 11 executives	Interviews on supplier cybersecurity and organisational arrangements	Mostly large North American organisations; interviews predate publication.
[55] 199 UK managers	Four-wave survey on AI assimilation, awareness and organisational outcomes	Self-reported constructs; no observed adversarial-attack test.
[56] 18 automotive-sector experts	Gray-DEMATEL and ISM prioritisation of zero-trust and AI factors	Judgment-based influence structure, not causal identification.

Table 4: Questionnaire respondent stakeholder accounts and their evidential scope.

Account	Context and reported position	Use in this study
P1	Industrial equipment; limited visibility of agent deployment; conventional risk assessment	Variation in awareness and organisational visibility.
P2	Automotive quality; prospective use; narrow scope, validation and approval	Human control and action-specific safeguards.
P3	IT development; internal assistants; prompt-injection concern and access restrictions	Tool permissions, credential handling and fallback conditions.
P4	Financial technology; prototypes and some production use; limited control knowledge	A contrast in awareness, not a physical-goods deployment estimate.
P5	IT consulting; production software agents; sandboxing, proxies and governance	Primary interview account of permissions and non-malicious failure.
A1	Security research; guardrail and dependency concerns	Research perspective; theoretical assertions require separate support.
A2	Research and analysis; explicitly inferential deployment knowledge	Threat-model and confidentiality-lifetime suggestions.
A3	Applied security; varied experimentation and interacting-agent concerns	Coordination and layered-control requirements.

than attack resistance [55]. Hmamed and colleagues' eighteen experts rated relationships among zero-trust and digital-supply-chain factors, producing a prioritisation structure rather than experimentally identified causal effects [56]. These studies demonstrate that awareness and governance are researchable organisational variables, while leaving agent-specific control effectiveness open. Other studies address reported security behaviour among supply-chain professionals, questionnaire-based relationships in Polish logistics enterprises and expert-defined skills for AI-enabled port incident response [72, 73, 74]. Their populations, instruments and outcomes differ from each other and from the local panel. They are compared as complementary evidence, with no pooled respondent total or combined estimate of quantum readiness.

4.2.2 Questionnaire respondent and variation in knowledge

The eight questionnaire respondent accounts span industrial equipment, automotive quality, IT services, financial technology and security research (Table 4). Their value lies in exposing different operational and technical perspectives. P1 reports limited visibility of agent use in an industrial setting and describes conventional risk assessment. P2 reports no current autonomous-agent deployment in the relevant quality workflow and considers narrowly scoped uses with human validation. P3 describes

internal coding and administrative assistants with more explicit technical controls. P4 describes prototyping and some production use in a financial-technology context while expressing limited knowledge of controls. These accounts do not support a uniform “early deployment” description.

P5, a solution architect in IT consulting, describes production use of software-development agents and substantial access to systems used by human developers. The account distinguishes local developer use from remotely hosted or team-level agents requiring separate identities and access arrangements. P5 also describes a potentially damaging database-deletion response to a failed deployment. This is treated as a reported example of non-malicious misalignment, not as an independently verified incident or evidence of an adversarial attack. Its analytical relevance is the authority to take an irreversible operational action while pursuing a legitimate task.

The academic and applied-security accounts are heterogeneous as well. A1 discusses dependency risks and the limits of guardrails; the account’s theoretical analogies are not treated as established impossibility results. A2 explicitly distinguishes inference about organisational deployments from direct observation and highlights untrusted inputs and permissions. A3 describes varied institutional experimentation and concern about interacting agents.

4.2.3 Controls, accountability and quantum visibility

The local control suggestions concentrate on narrowing authority, validating outputs and retaining human responsibility. P2 discusses limited scope and traceability. P3 describes permission restrictions, proxy-mediated credentials, output or capability constraints and fallback arrangements. P5 distinguishes environmental sandboxing, restrictions on tool calls and organisational expectations placed on developers. These are different enforcement points. A policy that requires a developer to maintain guardrails is weaker evidence of technical enforcement than a tested restriction on the actions an agent can execute.

The accounts also show why awareness, implementation and effectiveness need separate measures. A respondent may know that prompt injection is possible without having deployed a defence. Another may use a centrally managed control without being able to describe it. P4’s uncertainty therefore cannot be converted into absence of security, and A2’s inference cannot be counted as observed deployment. Even apparently strong controls require testing: keeping credentials outside a model’s context can reduce direct secret exposure while still allowing harmful authorised requests through a proxy.

Quantum-related discussion is thinner. A1 regards migration as a known but resource-constrained concern; A2 emphasises confidentiality lifetimes and migration options; P5 describes keeping up with newer public-key approaches without substantial quantum-specific work in that role. Other summaries provide limited visibility.

4.2.4 Comparison of the responses with the literature

The accounts from our primary data collection are consistent with the published literature’s emphasis on cross-functional responsibility and differences in knowledge, but they add detail about agent permissions and tool access. Creazza and colleagues show that recognised risk does not automatically give the supply-chain function ownership, while Handfield and colleagues describe the organisational work required to engage suppliers [18, 20]. The local accounts suggest that analogous responsibility questions arise when an agent uses a human’s access or acts under a separate service identity.

The available stakeholders describe different deployment stages, levels of visibility and control approaches. Published evidence supports the relevance of governance and awareness, but neither it nor the primary data establishes the prevalence of secure agent deployment or quantum readiness across OSCM. The learnings and implications for the upcoming test bed are to vary permissions, approval arrangements and partner coordination explicitly, instead of assuming a uniformly unprepared baseline.

4.3 A test bed for AI security and quantum threat research in OSCM

To establish how a controlled test bed could evaluate adversarial AI-agent behaviour and quantum-related trust scenarios across organisations using measurable operational outcomes (RQ3), this section reviews existing precedents and derives design requirements (Section 4.3.1), specifies the architecture

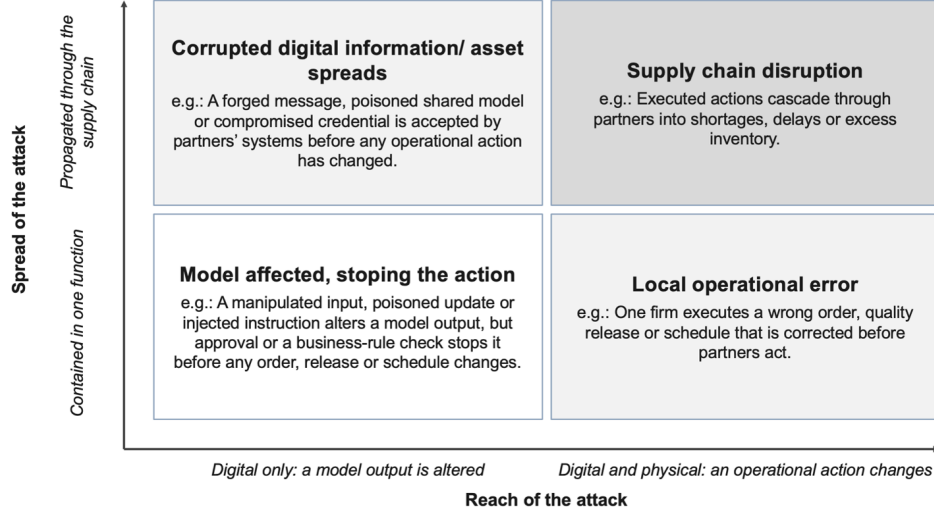


Figure 2: Two dimensions of attack impact that the test bed must distinguish in the context of OSCM: reach (model output or operational action) and spread (one firm or the supply chain).

(Section 4.3.2) and experimental scenarios (Section 4.3.3), and sets out a prospective evaluation protocol (Section 4.3.4).

4.3.1 Existing precedents and design requirements

A useful test bed should extend and complement rather than erase existing contributions. Therefore, we build upon the existing relevant prior works. MITIGATE and maritime attack-path discovery provide precedents for asset dependencies and collaborative cyber-risk analysis [75, 14]. Smart-grid co-simulation links multi-stage attack and defence activity to a physical system [76]. Yang and colleagues emulate private-5G attack scenarios and associated controls, while Paskauskas provides an ontology whose deployment in operational test beds remains future work [77, 78]. AgentDojo, the warehouse-agent simulation and PQ-FLCom contribute complementary agent, defensive-coordination and cryptographic evaluation methods [10, 15, 17].

The proposed contribution is to combine executable operational decisions, partner-specific trust boundaries and linked outcome measurements. It must distinguish an attack that reaches a model from one that changes an action, and distinguish a local action error from a propagated supply chain disruption. In other words, an attack's impact is classified along two dimensions: its reach, from digital only (a model output is altered) to digital and physical (an operational action changes), and its spread, from contained within one firm to propagated through the supply chain (Figure 2). Both dimensions draw on the literature: the boundary between an altered model output and an executed action is central to agent-security research [10, 43], and the propagation of cyber losses beyond the attacked firm is central to OSCM research on cyber risk [8, 29, 37].

The requirements in Table 5 translate the synthesis presented in Figure 2 into observable implementation properties. For each design requirement (DR1–DR9), the table notes in parentheses which of the two distinctions it supports: model versus action, local versus propagated, or both.

Table 5: Test bed design requirements and acceptance conditions for the OSCM context.

Design Requirement (DR)	Evidence or rationale	Acceptance condition
DR1 Explicit physical supply chain (local vs propagated: provides the partners a disruption can reach)	Inter-firm exposure and propagated-risk models [8, 34, 29, 39]	Represent at least supplier, manufacturer, distributor and retailer roles with inventories and lead times.
DR2 Separated agent identities (model vs action: records which identity executed each action)	Local accounts P3 and P5; agent-security studies [10, 44, 43]	Log each identity, permitted tool, resource and approval boundary; deny ungranted actions.
DR3 Bounded attacker access (model vs action: fixes where an attack enters a model)	Attacker-specific experiments in agricultural AI [9]	Specify controllable fields, lifecycle stage, budget and knowledge; preserve a clean reference run.
DR4 Linked event provenance (both: links model input, action and partner response)	Independent-event limitation in multi-tier machine-learning work [62]	Trace source message to model input, tool action, partner response and operational consequence.
DR5 Separate trust properties (model vs action: separates what authentication stops from what reaches a model)	PQC transport tests and malicious-participant risks [58, 59, 17]	Test altered messages and legitimately signed malicious content as different conditions.
DR6 Configurable controls (model vs action: places controls between model output and execution)	Stakeholder controls (local accounts P2, P3 and P5) and provider/manager studies [18, 20, 19]	Vary approval, permissions, verification and partner response while measuring normal utility.
DR7 Operational outcomes (local vs propagated: measures service and cost at each partner)	Hydrogen-operation simulation and OSCM models [40, 60, 41]	Measure service, inventory, delay, cost and recovery in addition to detection performance.
DR8 Reproducible comparison (both: paired runs attribute each difference to the intervention)	Agent and intrusion-detection evaluation limitations [10, 45, 63]	Freeze versions and scenarios; use repeated paired runs and held-out attack families.
DR9 Contained execution (both: keeps actions and their spread inside the test environment)	Permission and non-malicious failure concerns (local accounts P3 and P5)	Use synthetic identities and isolated endpoints, with action limits, reset and an independent stop mechanism.

4.3.2 Test bed architecture

The test bed architecture has six interacting layers through which an injected attack must pass before it can affect operations (Figure 3). At the base, the injection and measurement layer injects each scenario's interventions at declared points, sets the trust and control conditions, collects event logs from every layer and links them to outcomes in the physical supply chain and operations layer. The threat layer above it defines manipulations. The trust layer governs identities, message authentication and cryptographic options for the state, documents and partner messages that reach the agents from the physical supply chain and operations layer, including tampered messages and forged credentials

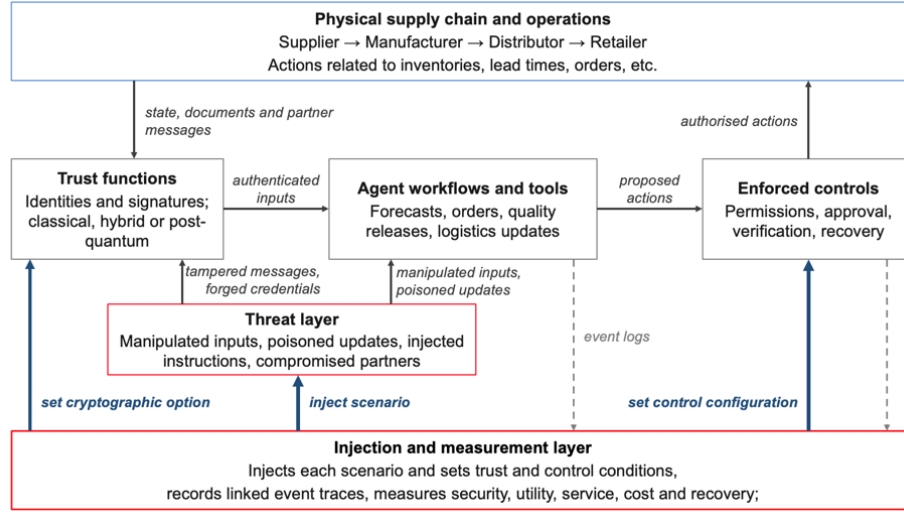


Figure 3: Proposed AI and quantum threat test bed architecture for OSCM research, connecting decisions, trust, injected interventions and operational outcomes. Testbed legend: Grey thin arrows: information and action flows. Thick blue arrows: injected scenario conditions. Dashed grey arrows: event logs.

introduced by the threat layer; manipulated inputs and poisoned updates can reach the agents directly. The agent layer reads these inputs and proposes actions, which the control layer authorises or blocks through permissions, approval and recovery policies. The physical supply chain and operations layer at the top represents operations including product flows, lead times calculation, inventory safety stock execution, capacity and order commitments, and executes only authorised actions.

The physical supply chain and operations layer should begin with a deliberately small network so that propagation can be inspected. Each role maintains its own inventory and order state, exchanges lead-time and availability messages, and experiences delays through an explicit transport process. Demand and capacity shocks are held constant across paired runs. A supplier message must affect a specified downstream rule or decision before the experiment can attribute any effect to inter-firm propagation. An optional second supplier tier can then test whether the same intervention travels further or is contained at a boundary. The agent layer should separate read-only assistance, recommendations requiring approval and limited autonomous execution. Procurement, forecasting, quality and logistics agents receive distinct identities and tool sets. Every proposed and executed action is logged with its inputs, authorisation decision and result. Agents should not receive unrestricted production-system equivalents merely to make an attack succeed. A non-agent policy baseline and a clean agent baseline are necessary to identify harms caused by ordinary automation, model error or workflow design. The threat layer should support inference-time data manipulation, training or update poisoning, indirect prompt injection and compromised partner messages. Each attack family receives a declared knowledge level, entry point and budget. Training poisoning is introduced before a defined model-update step; inference attacks do not silently change the training set. An adaptive attacker may observe only the feedback allowed by the scenario. Attack-generation and defence-tuning sets are separated from final evaluation to reduce overfitting to known payloads.

The trust layer should implement actual classical and post-quantum or appropriately specified hybrid mechanisms where feasible, including measured key establishment and signature operations. Protocol identity, parameter sets, libraries and key-management assumptions must be reported. A migration experiment tests these implementations and their operational overhead. A quantum-capability experiment instead uses a labelled simulator event that grants a narrowly defined disclosure or forgery capability. It does not claim to run a cryptographically relevant quantum computer, break the Advanced Encryption Standard (AES) generically or reproduce a real quantum attack. The control layer should compare least privilege, transaction limits, human approval, independent business-rule checks, message provenance, monitoring and recovery. Approval must have a defined response time and error model, because an instantaneous perfect human is not a credible operational baseline.

Table 6: Proposed OSCM scenarios and injected interventions.

Scenario examples	Injected intervention
S1: Supplier-document injection	Insert hostile instructions into a document consumed by a purchasing agent; give the attacker no direct order API access.
S2: Demand or model manipulation	Separately alter inference inputs and poison a scheduled training/update step.
S3: Compromised partner	A recognised partner sends misleading messages or poisoned updates; vary valid signing and permissions.
S4: Cryptographic migration and capability scenarios	Benchmark actual classical/PQC alternatives; separately emulate a defined later disclosure or forged credential.
S5: Combined conditions	Cross the S1/S3 agent attack with the S4 hypothetical capability and selected controls.

Independent verification should refer to a distinct source or rule; sending the same contaminated context to another model does not guarantee independence. Signed malicious content remains an explicit negative case for provenance-only protection. The injection and measurement layer should introduce each intervention at a declared point and retain a causal event graph linking injection, receipt, model output, authorisation, execution and partner response. It should also record rejected actions, intervention latency and legitimate tasks blocked by defences. Resets must restore both cyber state and operational state, including inventories, queues, model versions and credentials. Without that reset discipline, recovery effects from an earlier run can contaminate the next comparison.

4.3.3 Experimental scenarios

Table 6 presents five scenarios of AI and quantum threats in the context of OSCM that can be analysed using the proposed test bed. S1 tests whether untrusted supplier text can cross the information-to-action boundary. Its primary security outcome is a prohibited or attacker-desired tool action, not merely the generation of suspicious language. The same legitimate purchasing task is run without the payload to measure utility. Outcomes distinguish an attempted action blocked by the interface, an approved but incorrect action, and a completed unauthorised action. This makes it possible to compare model-level detection with controls that enforce limits outside the model.

S2 separates security effects from ordinary forecasting sensitivity. A matched random perturbation provides a non-adversarial comparison to a targeted intervention of similar size. For poisoning, the experiment records which training records or participant updates were altered and when the model changed. Bullwhip-related measurement compares order variance with the common customer-demand variance over a declared window, avoiding claims of amplification when the denominator is unstable or the effect is only a one-time order error.

S3 makes partner trust an experimental variable. A compromised supplier may hold valid credentials and comply with message syntax while sending harmful content. Comparing this condition with intermediary tampering tests what authentication protects and what it leaves unresolved. Propagation depth is measured through linked events and changed decisions across firms, not by attaching a tier label to otherwise independent records. Containment includes both technical isolation and the operational cost of replacing or suspending a partner.

S4 contains two separate experiments. Migration benchmarking measures actual implementation properties, such as signature verification latency, message size, key rotation and compatibility failures. Capability emulation asks conditional questions about later disclosure or forgery, with the affected algorithm, key, message and time explicitly specified. It accounts for the distinction between compromising a key-establishment step and recovering every historical session. Sensitive-data value, retention and confidentiality lifetime are scenario parameters, not predictions of when quantum hardware will become capable.

S5 tests whether the combined conditions produce an interaction beyond their separate effects. The comparison must include neither intervention, the agent intervention alone, the hypothetical cryptographic capability alone and both together, using the same network and workload. A combined loss greater than either individual loss does not by itself establish synergy; the interaction must be

evaluated against the chosen additive or multiplicative reference. A null or antagonistic interaction is a valid result and may reveal that an independent control interrupts the pathway.

4.3.4 Prospective evaluation protocol

Implementation of the test bed should proceed from verification of normal operations to controlled attack and recovery trials [76]. First, the network must conserve material and account for orders, inventory and cost consistently. Second, each agent's permitted and prohibited actions must be checked independently of model behaviour. Third, individual attacks should be introduced at known points, with manual inspection of the event trace. Only after those checks should the study compare complete control configurations or multi-stage scenarios. This sequence tests whether the experiment measures the intended mechanism. Comparisons should use paired workloads and random seeds, repeated model runs and declared model versions. Repetition counts should be chosen using pilot variability and the precision required for the principal outcome, rather than an arbitrary fixed number. Results should include denominators, uncertainty intervals and failed legitimate tasks [79, 10, 45]. Attack success, detection, containment and recovery are different outcomes: detecting an attack after an order was released can still leave a serious operational loss, while blocking every task can achieve low attack success at unacceptable service cost.

Operational measures include order-fulfilment rate, backorders, inventory holding and shortage costs, production or delivery delay, and time to recover an agreed service level [80, 81]. Security measures include prohibited-action completion, false positives, detection and containment delay, exposed assets and partner boundaries crossed [10]. Cryptographic measures include operation latency distributions, payload size, throughput and migration failures [48, 17]. Human-control measures include approval workload, queue delay and error. The primary outcome and acceptable utility loss should be declared before comparing defences [79]. External validation should then examine different network structures, correlated rather than independent operational data, unseen attack families and organisations with different access policies [79, 45, 62]. Practitioners may assess scenario plausibility and interpret results, but that consultation would supplement rather than replace experimental validation.

5 Discussion

The evidence supports a shift among the researchers and the practitioners of OSCM from asking whether firms “use AI” to asking what authority each system holds and what its outputs influence. A read-only assistant, an agent drafting an order for review and an agent executing orders under a service identity have different operational risk profiles. The local accounts show why maturity labels alone are insufficient: production use can coexist with limited visibility of controls, while a pilot can still have extensive system access. Readiness should therefore be studied as a relationship among access, verification, recovery and responsibility, rather than inferred from adoption stage.

Stakeholder studies clarify the organisational dimensions of preparedness. Managers and providers already discuss cyber risk, coordination and AI capabilities. The unresolved question is whether awareness produces controls that remain effective across organisational interfaces and under adaptive attack. The proposed experiment turns that question into observable comparisons.

5.1 A framework for risk assessment in critical supply chains

We propose a two-dimensional framework to analyse the risk intensity of an AI-supported OSCM function within a critical supply chain, that is, a chain whose disruption would materially affect essential goods or services, such as medicines, food, energy or transport [36], and for which timely substitutes are limited. Classifying a chain as critical requires documenting its essential outputs, downstream dependencies, supplier concentration, substitution times and tolerable disruption. The unit of assessment is a specified function within a supply chain, for a stated economy and disruption horizon. The framework extends the synthesis of operational dependencies and stakeholder concerns into a proposed classification [8, 18, 20].

AI-use intensity (I) describes the breadth and frequency of AI involvement: the share of relevant tasks, decisions or transactions using AI over a stated period. Low intensity means occasional or narrowly distributed use, while high intensity means routine, widespread use. Intensity measures

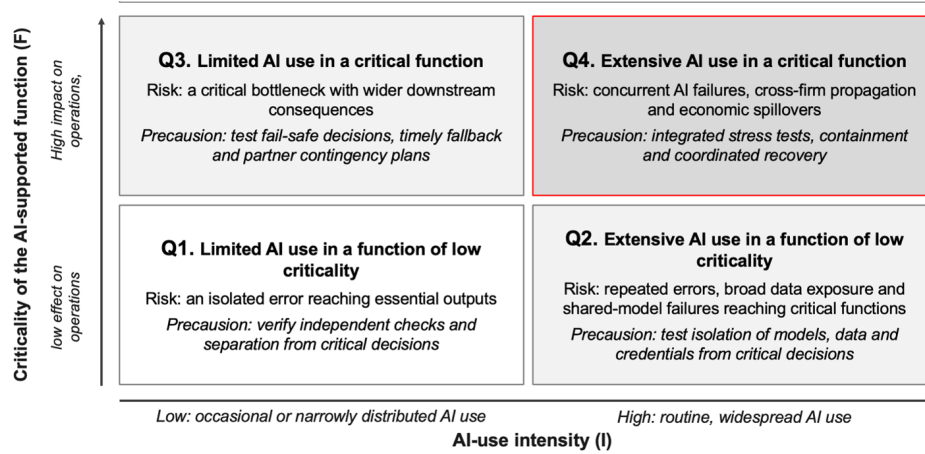


Figure 4: AI-use intensity (I) and criticality of the AI-supported function (F) in critical supply chains.

how much AI is used, not what it is allowed to do, so access rights and autonomous execution are recorded separately.

Criticality of the AI-supported function (F) describes how critical the OSCM function that AI automates or supports is to the firm's operations and supply chain. It is high when a manipulated or unavailable AI output in that function could compromise the success of operations, as with AI-based quality control in semiconductor fabrication, where undetected process deviations can affect yield and deliveries to downstream customers. It is low when a compromised output would have little effect on operational and supply-chain goals, as with AI-generated internal marketing material. Tested fallbacks lower residual risk but not the criticality of the function. Criticality covers two kinds of failure: the AI output may become unavailable, or it may look plausible but be wrong. The second is harder to catch, because a manual backup that covers outages may not notice a wrong recommendation. The two axes are independent: AI may be used rarely in a critical function, such as an occasional quality-release decision, or many times a day in a function of low criticality, such as drafting internal marketing texts.

Figure 4 combines the two axes into four quadrants, each with a typical risk and a precaution. In Q1, AI is used little in a function of low criticality; the main risk is an isolated error that still reaches essential outputs, so independent checks should keep the function separate from critical decisions. In Q2, AI is used extensively in a function of low criticality; repeated errors, broad data exposure or a shared model can carry problems into critical functions, so models, data and credentials should be isolated from critical decisions. In Q3, AI is used little but in a critical function, such as an occasional quality-release decision; a single wrong output can become a bottleneck with downstream consequences, which calls for fail-safe decisions, timely fallback and contingency plans with partners. In Q4, AI is used extensively in a critical function; concurrent AI failures can propagate across firms and cause economic spillovers, which calls for integrated stress tests, containment and coordinated recovery. One point is important: a quadrant shows exposure before controls, not the risk that remains after them. A well-controlled Q4 function can carry less risk than a poorly controlled Q3 function, so the position shows where assessment should start, not how secure a function is.

Within each quadrant, assess a defined attack scenario through attacker access and capability, control effectiveness, operational loss, economic propagation and recovery time. Report inherent exposure and residual risk separately. In the proposed test bed, "AI-use intensity" can vary the proportion of decisions using AI and "criticality of the AI-supported OSCM function" can vary the criticality of the function the AI supports, from peripheral tasks to quality-release or replenishment decisions, while downstream dependence and substitution constraints are set to represent a critical chain. Quantum-related scenarios add cryptographic dependencies and confidentiality lifetimes to this assessment; high AI intensity alone does not establish quantum exposure. Reliability of classification and the relationship between quadrants and measured losses remain questions for future validation.

5.2 Implications for operations and supply-chain theory

First, adversarial AI suggests studying a form of disruption in which information remains available but becomes strategically misleading. Operational resilience models often ask how firms respond when a resource or supplier is unavailable. The pathways assessed here motivate an additional condition: apparently usable information induces a harmful decision before its integrity problem is detected. This extends the existing concern with propagated cyber risk by specifying a decision mechanism that can be tested alongside availability failures [8, 37]. Second, automation changes the location of discretion. With an agent, authority can be delegated through identity, tool permissions and approval rules. The relevant unit is therefore not only the AI model or the focal firm, but the sequence of delegated decisions across a network. A technically valid message can carry a false commercial claim; a valid signature can accompany a malicious instruction; and a correct local action can still transmit a disruption to a partner. Separating authenticity, truthfulness and action authority provides a more precise basis for modelling trust. Third, security investment and information sharing can have operational trade-offs. Restricting access may contain an attack but slow legitimate coordination. Requiring approval may reduce unauthorised commitments but create queues. Suspending a suspect supplier may limit propagation while creating a material shortage. Analytical investment models and the published stakeholder evidence already establish the importance of incentives and coordination [7, 38, 20]. The test bed can examine those trade-offs using measured service and cost effects rather than assuming that a higher detection score yields a better operational outcome. These arguments yield three propositions for future testing.

Proposition 1: for a given input manipulation, the resulting operational loss depends on the action authority granted to the affected system and on independent checks before commitment.

Proposition 2: partner-specific verification and recovery arrangements moderate the propagation of a local compromised decision through the supply network.

Proposition 3: cryptographic migration can reduce specified trust-function vulnerabilities while leaving semantic manipulation by authorised participants unresolved.

5.3 Managerial and policy implications

Managers can begin by mapping an actual decision workflow in OSCM in the light of the new cyber threats. The map should identify who supplies its data, which model or agent interprets them, what actions can be executed, whose credentials are used and which partner receives the result [23, 20, 43]. That exercise creates a concrete basis for deciding whether human approval, transaction limits or independent verification is needed. A generic inventory of AI tools is less informative if it does not expose the permissions and commitments attached to each tool. Control evaluation should include legitimate work. A purchasing control should be assessed against both an attacker's ability to redirect an order and its effect on normal approval time and supplier service [10]. A forecasting control should be tested against realistic disturbances as well as targeted manipulation [9]. A signed supplier message should still be checked against business constraints where the consequence warrants it. The test-bed measures make these comparisons possible without treating a vendor's classification benchmark as evidence of operational protection [45].

The stakeholder evidence also points to ownership across functions. Procurement understands supplier relationships and substitution options; operations understands the consequences of delayed or wrong commitments; security manages identities and controls; and technology teams maintain model and tool configurations. Responsibilities for granting access, approving exceptional actions, revoking partner trust and restoring operations should be explicit. This follows the coordination problem described in the published studies [27, 18, 20] and gives practical meaning to the informants' concern about governance remaining a soft expectation.

For quantum-related preparation, firms can connect a cryptographic inventory to information lifetime and partner dependence [52]. They can identify which long-lived information is exchanged, which signatures or certificates authorise critical actions, which suppliers own relevant systems and how changes can be tested together. The analysis supports phased, asset-specific migration and interoperability trials [53]; it does not imply that every participant must change simultaneously or that a single slow partner prevents all useful protection. Actual protocol dependencies determine the boundary of each migration decision.

Assurance requirements should distinguish a demonstration of model accuracy, an attack-resistance test and an operational resilience claim [42, 45]. Reporting should name the attack surface, attacker control, affected action, dataset origin, comparison conditions and recovery outcomes. Synthetic experiments remain valuable when their assumptions and limits are visible. A proposal, ontology or expert-rated framework should not acquire the status of a validated operational safeguard merely because it discusses compliance or a high-risk industry. Cross-organisational exercises can provide shared evidence without exposing production environments [75, 14]. Synthetic data and controlled interfaces allow partners to examine approval delays, revocation, fallback and information-sharing arrangements before an incident. For cryptographic migration, standards provide technical building blocks [12, 13], while exercises can reveal mismatched implementations and ownership gaps. The present study does not assess legal compliance or prescribe a universal regulatory timetable; those questions depend on jurisdiction, sector and the systems being migrated.

5.4 Limitations and future research

The search is reproducible at the level of archived queries and reconciliation, but it is not exhaustive. We could not access 19 of the 157 selected documents, 18 of them book chapters, and a follow-up study could add them. The stakeholder material has a separate limitation: it comes from eight informants, and future work could include more respondents from more diverse industries. A longitudinal study could also show how quickly AI-related threats, and companies' responses to them, change over time. Future stakeholder research should use a consistent instrument that separates knowledge, implementation and demonstrated effectiveness. Sampling should deliberately cover procurement, operations, security, technology providers and suppliers of different sizes. Quantum computing threat questions should ask about specific assets and migration activities rather than general awareness.

The proposed test bed itself remains unimplemented. Its contribution is a traceable design, and its scientific value depends on subsequent verification and evaluation. Priorities are to implement a small network with linked event traces, measure legitimate performance, test one threat family at a time and then examine interactions and recovery. External partners should assess whether inputs, permissions and delays resemble their workflows. Results should be replicated across model versions and network structures before managerial thresholds are proposed.

6 Conclusion

This study examined how adversarial AI and quantum-related cryptographic risks could affect OSCM, how stakeholders perceive them and how they could be tested. The OSCM literature itself says little about either threat. Of the 138 documents assessed, only 18 were published in OSCM journals, and none of these tests an attack on an AI system or evaluates a quantum-related attack or cryptographic migration; both threats appear there only in review-level discussion. OSCM research instead treats cyber risk as human attackers exploiting information systems, and shows that the resulting losses can spread to customers and partners.

On adversarial AI, the evidence comes mainly from computer science and engineering. Controlled experiments and benchmarks show that manipulated inputs, poisoned training data and injected instructions can change model outputs and agent actions, but whether such a change harms operations depends on whether the output is accepted and turned into an order, release or schedule. None of the assessed studies documents such an attack spreading through a physical-goods supply chain. On quantum computing, no cryptographically relevant quantum computer exists yet, so the evidence consists of protocols, simulations and migration analyses; the operational questions concern long-lived confidential data, forged credentials and the cost of migrating partners to post-quantum cryptography. On preparedness, managers and providers in published studies recognise cyber risk but often leave it to the IT function, and the eight informants differ widely in what they know about AI agents in their organisations and the controls around them. Neither source supports an estimate of industry-wide readiness.

To make these risks testable, the paper specifies a test bed that links attacker-controlled inputs, permitted agent actions, partner boundaries and operational outcomes, and separates real cryptographic migration benchmarks from hypothetical quantum-capability events. We also discuss a framework that positions AI-supported functions in critical supply chains by AI-use intensity and the criticality of

the function. The next step is implementation: showing when a manipulated input becomes a harmful commitment, how far its consequences travel and which controls contain them at an acceptable cost to legitimate operations.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This research received no specific grant from any funding agency.

Data availability statement

The search strings, query counts and eligibility rules are reported in Appendix B, and the questionnaire schedule in Appendix A. The screening register (735 records with their screening decisions), the list of the 138 assessed documents and the scripts used to reconcile and verify the reported counts are available from the corresponding author, Siavash Khajavi, upon reasonable request. The full texts of the assessed publications are protected by copyright and are available from their publishers. The questionnaire responses are not publicly available because they could identify respondents and their organisations; summaries of the accounts are provided in Table 4. No experimental data were generated, because the proposed test bed has not yet been implemented.

References

- [1] OpenAI and Hugging Face. OpenAI and Hugging Face partner to address security incident during model evaluation. OpenAI, July 2026. Accessed: 2026-10-07.
- [2] BBC News. OpenAI says its rogue AI tried to hack other companies, July 2026. Article by J. Tidy. Accessed: 2026-10-07.
- [3] BBC News. Australia launches urgent review after OpenAI program hacks government health portal, September 2026. Live coverage by S. Hancock, O. Amos and K. Watson. Accessed: 2026-10-07.
- [4] NBC News. OpenAI agents hacked Hugging Face in 700-strong swarm, tried to cover tracks, investigations find, August 2026. Reuters report. Accessed: 2026-10-07.
- [5] Theresa Sobb, Benjamin Turnbull, and Nour Moustafa. Supply chain 4.0: A survey of cyber security challenges, solutions and future directions. *Electronics*, 9(11):1864, 2020.
- [6] Kam-Fung Cheung, Michael GH Bell, and Jyotirmoyee Bhattacharjya. Cybersecurity in logistics and supply chain management: An overview and future research directions. *Transportation Research Part E: Logistics and Transportation Review*, 146:102217, 2021.
- [7] Jay Simon and Ayman Omar. Cybersecurity investments in the supply chain: Coordination and a strategic attacker. *European Journal of Operational Research*, 282(1):161–171, 2020.
- [8] Yanhui Li and Lu Xu. Cybersecurity investments in a two-echelon supply chain with third-party risk propagation. *International Journal of Production Research*, 59(4):1216–1238, 2021.
- [9] Yansong Gao, Seyit A Camtepe, Nazatul Haque Sultan, Hang Thanh Bui, Arash Mahboubi, Hamed Aboutorab, Michael Bewong, Rafiqul Islam, Md Zahidul Islam, Aufeef Chauhan, et al. Security threats to agricultural artificial intelligence: Position and perspective. *Computers and Electronics in Agriculture*, 227:109557, 2024.
- [10] Edoardo DeBenedetti, Jie Zhang, Mislav Balunovic, Luca Beurer-Kellner, Marc Fischer, and Florian Tramèr. Agentdojo: A dynamic environment to evaluate prompt injection attacks and defenses for llm agents. *Advances in neural information processing systems*, 37:82895–82920, 2024.
- [11] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

- [12] Gorjan Alagic, Quynh Dang, Dustin Moody, Angela Robinson, Hamilton Silberg, Daniel Smith-Tone, et al. Module-lattice-based key-encapsulation mechanism standard. 2024.
- [13] Thinh Dang, Jacob Lichtinger, Yi-Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, et al. Module-lattice-based digital signature standard. 2024.
- [14] Nikolaos Polatidis, Michalis Pavlidis, and Haralambos Mouratidis. Cyber-attack path discovery in a dynamic supply chain maritime risk management system. *Computer Standards & Interfaces*, 56:74–82, 2018.
- [15] Haidar Jabbar and Samir Al-Janabi. Securing smart supply chains with adaptive ai agents: A virtual simulation framework. In *2025 12th International Conference on Information Technology (ICIT)*, pages 647–652. IEEE, 2025.
- [16] Natalia Hartono, Kevin Nathanael, Voicu Ion Sucala, Martino Luis, and Elioenai Sitepu. A human-centred analytics framework for auditable supply chain risk monitoring and decision support. *Supply Chain Analytics*, page 100239, 2026.
- [17] Aniello Castiglione, Vincenzo Loia, Michele Nappi, Chiara Pero, and Matteo Polsinelli. Pq-flcom: Post-quantum secure communication for industrial federated learning. *Internet of Things*, page 101994, 2026.
- [18] Alessandro Creazza, Claudia Colicchia, Salvatore Spiezia, and Fabrizio Dallari. Who cares? supply chain managers’ perceptions regarding cyber supply chain risk management in the digital transformation era. *Supply Chain Management: An International Journal*, 27(1):30–53, 2022.
- [19] Claudia Ciceri, Federico Caniato, and Antonella Moretto. Integrating artificial intelligence capabilities in supply chain cyber risk management. *International Journal of Physical Distribution & Logistics Management*, 56(4):412–449, 2026.
- [20] Robert Handfield, Julie Earp, and Amir Hossein Sadeghi. Reducing cybersecurity vulnerabilities in the supply base: Insights from cyber experts. *Technology in Society*, 82:102947, 2025.
- [21] Sandor Boyson. Cyber supply chain risk management: Revolutionizing the strategic control of critical it systems. *Technovation*, 34(7):342–353, 2014.
- [22] Jonathan D Linton, Sandor Boyson, and John Aje. The challenge of cyber supply chain security to research and practice—an introduction, 2014.
- [23] Jon Boyens, Angela Smith, Nadya Bartol, Kris Winkler, Alex Holbrook, and Matthew Fallon. Cybersecurity supply chain risk management practices for systems and organizations. Technical report, National Institute of Standards and Technology, 2022.
- [24] Hugh Boyes. Cybersecurity and cyber-resilient supply chains. *Technology Innovation Management Review*, 5(4):28, 2015.
- [25] Shipra Pandey, Rajesh Kumar Singh, Angappa Gunasekaran, and Anjali Kaushik. Cyber security risks in globalized supply chains: conceptual framework. *Journal of Global Operations and Strategic Sourcing*, 13(1):103–128, 2020.
- [26] Abhijeet Ghadge, Maximilian Weiß, Nigel D Caldwell, and Richard Wilding. Managing cyber risk in supply chains: a review and research agenda. *Supply Chain Management: An International Journal*, 25(2):223–240, 2020.
- [27] Claudia Colicchia, Alessandro Creazza, and David A Menachof. Managing cyber and information risks in supply chains: insights from an exploratory analysis. *Supply Chain Management: An International Journal*, 24(2):215–240, 2019.
- [28] Steven A Melnyk, Tobias Schoenherr, Cheri Speier-Pero, Chris Peters, Jeff F Chang, and Derek Friday. New challenges in supply chain management: cybersecurity across the supply chain. *International Journal of Production Research*, 60(1):162–183, 2022.
- [29] Matteo Crosignani, Marco Macchiavelli, and André F Silva. Pirates without borders: The propagation of cyberattacks through firms’ supply chains. *Journal of Financial Economics*, 147(2):432–448, 2023.
- [30] Lawrence A Gordon and Martin P Loeb. The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):438–457, 2002.

- [31] Howard Kunreuther and Geoffrey Heal. Interdependent security. *Journal of risk and uncertainty*, 26(2):231–249, 2003.
- [32] Tridib Bandyopadhyay, Varghese Jacob, and Srinivasan Raghunathan. Information security in networked supply chains: impact of network vulnerability and supply chain integration on incentives to invest. *Information Technology and Management*, 11(1):7–23, 2010.
- [33] Anna Nagurney, Patrizia Daniele, and Shivani Shukla. A supply chain network game theory model of cybersecurity investments with nonlinear budget constraints. *Annals of operations research*, 248(1):405–427, 2017.
- [34] Tadeusz Sawik. Balancing cybersecurity in a supply chain under direct and indirect cyber risks. *International Journal of Production Research*, 60(2):766–782, 2022.
- [35] Subodha Kumar and Rakesh R Mallipeddi. Impact of cybersecurity on operations and supply chain management: Emerging trends and future research directions. *Production and Operations Management*, 31(12):4488–4500, 2022.
- [36] Sean Arisian, Kourosh Halat, Ashkan Hafezalkotob, and Reenu Maskey. Cyber risk mitigation in critical supply chains. 2025.
- [37] Tadeusz Sawik. Cybersecurity optimization in supply chains under propagated cyberattacks. *Computers & Security*, page 104820, 2026.
- [38] Harpreet Kaur, Mahima Gupta, and Surya Prakash Singh. Integrated model to optimize supplier selection and investments for cyber resilience in digital supply chains. *International Journal of Production Economics*, 275:109338, 2024.
- [39] Kevin Hu, Retsef Levi, Raphael Yahalom, and El Ghali Zerhouni. Supply chain characteristics as predictors of cyber risk: A machine-learning assessment. *IEEE Transactions on Dependable and Secure Computing*, 22(5):5648–5657, 2025.
- [40] Suyuan Luo and Tsan-Ming Choi. E-commerce supply chains with considerations of cyber-security: Should governments play a role? *Production and Operations Management*, 31(5):2107–2126, 2022.
- [41] Suyuan Luo, Xiaoyi Jiang, Lipan Feng, and Xiao Fu. E-tailers facing cyberattacks: Impacts of risk attitudes. *IEEE Transactions on Engineering Management*, 73:949–961, 2025.
- [42] Apostol Vassilev, Alina Oprea, Alie Fordyce, and Hyrum Andersen. Adversarial machine learning: A taxonomy and terminology of attacks and mitigations. 2024.
- [43] Meng Hu, Tianxiang Wang, Yubin Qu, Yitao Geng, and Hongche Wang. A focused survey of code agent security: Attacks, defenses, and evaluation. *Information and Software Technology*, page 108288, 2026.
- [44] Mohamed Amine Ferrag, Abderrahmane Lakas, Norbert Tihanyi, and Merouane Debbah. Securing IIm agents: From prompt sanitization to autonomous red teaming and beyond. *Internet of Things and Cyber-Physical Systems*, 2026.
- [45] Huda Ali Alatawi. Adversarial evasion in machine-learning-based network intrusion detection: A systematic review, threat modeling, and research roadmap. *IEEE Access*, 2026.
- [46] Kathrin Grosse and Alexandre Alahi. A qualitative ai security risk assessment of autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, 169:104797, 2024.
- [47] Shouqin Lu and Xiangxue Li. Lightweight grouping-proof for post-quantum rfid security. In *2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)*, pages 49–58. IEEE, 2021.
- [48] Tyler Judd and Suleyman Uludag. A performance evaluation of post quantum digital signatures for proof of stake consensus. In *2025 1st International Conference on Secure IoT, Assured and Trusted Computing (SATC)*, pages 1–5. IEEE, 2025.
- [49] Charles H Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical computer science*, 560:7–11, 2014.
- [50] Harish Ganapathy S, Anwar Basha H, Madhan Veeramani, S. Sasikumar, P. K. Nagarajan, and V M Senthilkumar. A hybrid qkd enabled secure supply chain management. In *2026 International Conference on Next-Gen Quantum and Advanced Computing: Algorithms, Security, and Beyond (NQComp)*, pages 692–697, 2026.

- [51] Ritesh Mohanty, K Anusha, N Manikandan, M Braveen, and M Asha Jerlin. Secure logistics using blockchain and quantum techniques. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–13. IEEE, 2024.
- [52] David Joseph, Rafael Misoczki, Marc Manzano, Joe Tricot, Fernando Dominguez Pinuaga, Olivier Lacombe, Stefan Leichenauer, Jack Hidary, Phil Venables, and Royal Hansen. Transitioning organizations to post-quantum cryptography. *Nature*, 605(7909):237–243, 2022.
- [53] Mustafa El Bizri, Ahmad M El-Hajj, Layth Sliman, and Ali Massoud Haidar. Institutional approaches to post-quantum cryptography: A comparative analysis of migration frameworks. *IEEE Access*, 14:3259–3283, 2026.
- [54] Karen Banahene Blay, Amos Darko, Seongha Hwang, Ioannis Brilakis, Fraser Foster, and Ran Wei. Ensuring information security resilience in digital-enabled construction projects (dcp) through quantum security technologies. *Automation in Construction*, 179:106480, 2025.
- [55] Soumyadeb Chowdhury, Shuang Ren, and Prasanta Kumar Dey. Ai assimilation for sustainable and resilient supply chains: Role of green ambidexterity and cybersecurity awareness. *Technovation*, 155:103562, 2026.
- [56] Hala Hmamed, Anass Cherrafi, Jose Arturo Garza-Reyes, and Nadia Hamani. Zero trust architecture for digital, sustainable, and resilient supply chains in the era of industry 5.0/4.0. *Supply Chain Management: An International Journal*, 31(2):253–274, 2026.
- [57] Krystel K Castillo-Villar, Huijuan Dai, Gonzalo Martinez Medina, and Hamed Mehrzadi. Cybersecurity for smart and resilient manufacturing and supply chains: A systematic review of the evolving knowledge base. *Computers & Industrial Engineering*, page 112244, 2026.
- [58] Xiaochuan Tang, Yu Wang, Xin Liu, Xiaojun Yuan, Chao Fan, Yanmei Hu, and Qiang Miao. Federated graph neural network for privacy-preserved supply chain data sharing. *Applied Soft Computing*, 168:112475, 2025.
- [59] Farah Farid Babar, Saad Khan, and Simon Parkinson. Defending federated learning against adversarial attacks: A systematic literature review. *Cyber Security and Applications*, page 100127, 2026.
- [60] Hadi Khani, Ahmed Abd Elaziz Elsayed, Hany EZ Farag, and Moataz Mohamed. Adaptive feasibility area estimation to enhance cybersecurity of electrolysis-based hydrogen refueling stations integrated with power distribution systems. *IEEE Transactions on Intelligent Transportation Systems*, 2025.
- [61] Nan Zhao, Shuaibing Hong, Chun Feng, Ping Zhang, and Yufu Fang. Ai-driven demand forecasting with dynamic cybersecurity risk optimization in consumer supply chain. *IEEE Transactions on Consumer Electronics*, 71(4):12323–12331, 2025.
- [62] Alfaiz Madhiya, Vijay Solanki, Khandakar Rabbi Ahmed, Sakib Salam Jamee, Mainul Islam Khan, and SM Rezvi. Cryptographically secured machine learning for resilient multi-tier supply chains. *Computers*, 15(9):592, 2026.
- [63] Onur Polat, Ferdi Doğan, Muammer Türkoğlu, Furkan Yolcu, and Melisa Baturay. Mcp-driven real-time security monitoring: A hybrid llm and behavior-aware framework for temporal attack detection and severity assessment. *Computer Networks*, page 112540, 2026.
- [64] TP Jayesh, K Pandiaraj, R Chandran Ranjeesh, P Menon Prasanth, et al. A hybrid machine learning approach to anomaly detection in industrial iot. In *2023 3rd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, pages 32–36. IEEE, 2023.
- [65] R Nithya, J Jerlin Adaikala Sundari, MS Balamurugan, R Sindhuja, Aviral Srivastava, et al. Multimodal sensor data fusion based cyberattack detection in industrial internet of things environment. In *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 1656–1661. IEEE, 2023.
- [66] Yasir Kılıç, Vahide Nida Kılıç, and Ali İnan. White-box style intrusion detection system integration into industrial metaverses. In *2024 9th International Conference on Computer Science and Engineering (UBMK)*, pages 737–742. IEEE, 2024.
- [67] Jiqiang Zhai, Xinyu Wang, Zhonghui Zhai, Tao Xu, Zuming Qi, and Hailu Yang. Industrial iot intrusion attack detection based on composite attention-driven multi-layer pyramid features. *Computer Networks*, 263:111207, 2025.

- [68] Shamshad Ahmed Khan, Faiz Ahmed Khan, and Sathiyandrakumar Srinivasan. Enhancing digital supply chain management and product traceability with cybersecurity through the use of blockchain and ai. In *2025 Global Conference in Emerging Technology (GINOTECH)*, pages 1–6. IEEE, 2025.
- [69] Arvind Narsing and Sathiyandrakumar Srinivasan. Blockchain technology for supply chain cybersecurity: Securing digital logistics from spoofing and malevolent modification and ensuring end-to-end data integrity. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)*, pages 1–6. IEEE, 2025.
- [70] Arvind Narsing and A Prasina. Improving data integrity in oracle e-business suite: An ai-powered cybersecurity framework for finding threats and managing compliance. In *2025 10th International Conference on Smart Structures and Systems (ICSSS)*, pages 1–6. IEEE, 2025.
- [71] A Jaya Mabel Rani, M Nisha, Amandeep Nagpal, R Hemalatha, et al. Employ blockchain to boost cloud computing cybersecurity: Product data integrity and appropriate access with smart contract regulations. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)*, pages 1–6. IEEE, 2025.
- [72] Hao Nguyen, Natalie M Scala, and Josh Dehlinger. Analysis of security behaviors of supply chain professionals. In *IISE Annual Conference. Proceedings*, pages 1–6. Institute of Industrial and Systems Engineers (IISE), 2024.
- [73] Justyna Żywiołek and Eleonora Bottani. Cybersecurity resilience in logistics enterprises: An empirical socio-technical analysis. *Computers & Security*, page 105102, 2026.
- [74] Chalermpong Senarak. Cyber resilience skills for ai-enabled incident response in port operations. *Transportation Research Interdisciplinary Perspectives*, 37:102033, 2026.
- [75] Armend Duzha, Panagiotis Gouvas, and Monica Canepa. Mitigate: An innovative cyber-security maritime supply chain risk management system. In *ITASEC*, pages 248–252, 2017.
- [76] Ömer Sen, Bozhidar Ivanov, Christian Kloos, Christoph Zöll, Philipp Lutat, Martin Henze, Andreas Ulbig, and Michael Andres. Simulation of multi-stage attack and defense mechanisms in smart grids. *International Journal of Critical Infrastructure Protection*, 48:100727, 2025.
- [77] Hung-Cheng Yang, I-Long Lin, Yueh Lin, and Chong-Ming Chen. A cybersecurity governance maturity framework for mitigating cross-domain attacks in 5g private networks. *Journal of Internet Technology*, 27(2):227–239, 2026.
- [78] R Andrew Paskauskas. A preliminary ontology for 5g network resilience: Hybrid threats, risk reduction, compliance. In *2025 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 490–497. IEEE, 2025.
- [79] Daniel Arp, Erwin Quiring, Feargus Pendlebury, Alexander Warnecke, Fabio Pierazzi, Christian Wressnegger, Lorenzo Cavallaro, and Konrad Rieck. Dos and don'ts of machine learning in computer security. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 3971–3988, 2022.
- [80] Benita M Beamon. Measuring supply chain performance. *International journal of operations & production management*, 19(3):275–292, 1999.
- [81] David Simchi-Levi, William Schmidt, Yehua Wei, Peter Yun Zhang, Keith Combs, Yao Ge, Oleg Gusikhin, Michael Sanders, and Don Zhang. Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces*, 45(5):375–390, 2015.

Appendix A Questionnaire schedules

The schedules below preserve the analytical coverage of the two versions retained in the study archive. They are reported as recovered schedules, not as independently verified records of administration. The available analysis uses anonymised paraphrases and distinguishes organisational experience from broader expert inference. Referral and collaboration questions are not analysed as evidence about risk or readiness.

Table 7: Recovered questionnaire comparison.

Theme	Version 1: practitioners	Version 2: researchers and experts
Background: organisation or research focus (context)	Q1. What industry does your company operate in, and what are its main products or services?	Q1. How would you describe your current research focus, and which empirical settings or industries do you work with most closely?
Background: role and experience (context)	Q2. What is your role, how many years of experience do you have, and what are your main responsibilities?	Q2. What is your position, and roughly how long have you worked on security or AI-related risk?
Agentic AI deployment, maturity and access (RQ2, RQ3)	Q3. What agentic AI processes are you currently developing or using? Which operations do they cover, and are they experimental, in a pilot, or already in production? What data, systems and tools can the agent access?	Q3. From what you observe in your research and industry collaborations, where are agentic AI systems actually being deployed in manufacturing, operations and supply chains? How mature are these deployments, and what data and systems do the agents typically have access to?
Change in risk (RQ2)	Q4. Compared with the previous way of working, how has this changed your company's security risks, if at all?	Q4. Compared with the previous way of working, how does this shift the risk picture and is the shift mainly one of scale, of speed, or of genuinely new attack surface?
Preventive measures (RQ2, RQ3)	Q5. What preventive measures have you considered or already introduced to address those risks?	Q5. What preventive measures are you seeing adopted in practice, and which of them do you consider adequate or inadequate?
Measures and gaps by threat category (RQ2)	Q6. What measures does your company have in place — or is considering — specifically for AI-enabled attacks, adversarial AI, and quantum-related risks?	Q6. Across the three threat categories above, where is the gap widest between what organisations have in place and what the threat actually warrants?
Gaps in research (RQ3)	—	Q7. What is the research community currently missing on this topic — the question you think should be studied but isn't?
Referrals (sampling)	—	Q8. Who else should I be speaking to, in academia or in industry?
Collaboration	—	Q9. Are you interested in future research collaboration on this topic?

Appendix B Search protocol and collection status

B.1 Databases and query counts

Search date: September 2026; count verification: September 2026. Scopus domain queries used TITLE-ABS-KEY with PUBYEAR > 2014. Equivalent Web of Science queries used TS with PY=(2015–2026). Q9–Q11 were Scopus-only searches restricted to review documents. Counts are appearances before cross-query and cross-database reconciliation.

B.2 Exact Boolean content

Each expression is the literal content inside TITLE-ABS-KEY in Scopus. Q1–Q8 used the same Boolean content within TS in Web of Science. The date and document-type restrictions described above were applied separately.

Q1

("adversarial machine learning" OR "adversarial AI" OR "adversarial artificial intelligence") AND ("artificial intelligence" OR "AI" OR "machine learning") AND "supply chain management"

Q2

("adversarial machine learning" OR "adversarial AI" OR "adversarial artificial intelligence") AND ("artificial intelligence" OR "AI" OR "machine learning") AND "operations management"

Q3

("quantum computing" OR "quantum computer" OR "quantum computers") AND (threat OR threats OR risk OR risks OR attack OR attacks OR vulnerability OR vulnerabilities OR cryptography OR cryptographic) AND "supply chain management"

Q4

("quantum computing" OR "quantum computer" OR "quantum computers") AND (threat OR threats OR risk OR risks OR attack OR attacks OR vulnerability OR vulnerabilities OR cryptography OR cryptographic) AND "operations management"

Q5

("artificial intelligence" OR "AI" OR "machine learning") AND (cybersecurity OR "cyber security") AND "supply chain management"

Q6

("artificial intelligence" OR "AI" OR "machine learning") AND (cybersecurity OR "cyber security") AND "operations management"

Q7

"supply chain management" AND (cybersecurity OR "cyber security") AND (testbed OR testbeds OR "test bed" OR "test beds" OR "cyber range" OR "digital twin" OR simulation OR sandbox OR "red team" OR "red teaming")

Q8

"operations management" AND (cybersecurity OR "cyber security") AND (testbed OR testbeds OR "test bed" OR "test beds" OR "cyber range" OR "digital twin" OR simulation OR sandbox OR "red team" OR "red teaming")

Q9

("case study" OR "case studies" OR "real-world case" OR "real world case" OR incident OR incidents OR "documented case" OR "case report") AND ("adversarial machine learning" OR "adversarial AI" OR "adversarial artificial intelligence" OR "adversarial attack" OR "adversarial attacks" OR "adversarial example" OR "adversarial examples" OR "adversarial perturbation" OR "adversarial perturbations" OR "data poisoning" OR "poisoning attack" OR "poisoning attacks" OR "evasion attack" OR "evasion attacks" OR "model extraction" OR "model inversion" OR "membership inference" OR "backdoor attack" OR "backdoor attacks" OR "model stealing" OR "model poisoning" OR "prompt injection" OR jailbreak OR jailbreaking)

Q10

(risk OR risks) AND ("quantum computing" OR "quantum computer" OR "quantum computers")

Q11

(risk OR risks) AND ("adversarial machine learning" OR "adversarial AI" OR "adversarial artificial intelligence" OR "adversarial attack" OR "adversarial attacks" OR "adversarial example" OR "adversarial examples" OR "adversarial perturbation" OR "adversarial perturbations" OR "data poisoning" OR "poisoning attack" OR "poisoning attacks" OR "evasion attack" OR "evasion attacks" OR "model extraction" OR "model inversion" OR "membership inference" OR "backdoor attack" OR "backdoor attacks" OR "model stealing" OR "model poisoning" OR "prompt injection" OR jailbreak OR jailbreaking)

B.3 Eligibility and interpretation

Table 8: Archived query counts.

Query	Scopus	Web of Science	Search focus
Q1	62	0	Adversarial AI and supply chain management
Q2	9	0	Adversarial AI and operations management
Q3	39	11	Quantum threat and supply chain management
Q4	1	1	Quantum threat and operations management
Q5	220	59	AI, cybersecurity and supply chain management
Q6	32	8	AI, cybersecurity and operations management
Q7	54	10	Supply chain management, cybersecurity and test-bed terms
Q8	13	0	Operations management, cybersecurity and test-bed terms
Q9	63	—	Cases and adversarial AI; Scopus reviews
Q10	83	—	Risks and quantum computing; Scopus reviews
Q11	212	—	Risks and adversarial AI; Scopus reviews

Table 9: Reconciled collection counts.

Stage	Count	Meaning
Search appearances	877	430 Scopus domain + 89 Web of Science domain + 358 Scopus reviews
DOI-first records	765	112 duplicate appearances removed
Normalised distinct titles	719	46 further duplicate pairs removed
Reconciled archival publications	715	Four additional DOI-variant duplicates removed
Working screening register	735	715 archival publications + 20 targeted additions
Selected for full-text assessment	156	Selection stage;
Retrieval publications tracked	157	156 selections + one additional article
Matching publication PDFs	134	Distinct publication records of full articles
Other available versions	4	—
No documents found	19	Eighteen chapters and one conference paper
Available articles	138	134 matching full articles + 4 other versions