Agentic AI for Supply Chain Resilience:

A Mixed-Method Study of Agentic AI Design Principles Informed by Avionics Practices

Alberto Pogna

Delft University of Technology

TU Delft

Agentic AI for Supply Chain Resilience:

A Mixed-Method Study of Agentic AI Design Principles Informed by Avionics Practices

by

Alberto Pogna

Alberto
Pogna

Chair and first supervisor:	Patrick Stokkink
Second supervisor:	Arjan van Binsbergen
Project Duration:	September, 2025 - March, 2026
MSc:	Transport, Infrastructure and Logistics

TU Delft

Summary

Global supply chains are increasingly exposed to disruptions arising from climate events, geopolitical shifts, and cyber threats. Artificial intelligence (AI) has the potential to enhance resilience but may also introduce new risks if not effectively governed. This thesis examines organisational strategies for designing and governing AI to strengthen supply chain resilience while maintaining manageable complexity. A mixed-methods approach was adopted, structured around a two-iteration Double Diamond process. The first iteration involved semi-structured interviews with three AI experts and three supply chain resilience experts, followed by a focus group in which practitioners collaboratively translated interview findings into concrete agent concepts and operational governance boundaries, as well as a case study of a specific AI implementation. The second iteration expanded the analysis through targeted validation interviews, a cross-domain analogy with the aviation industry, and a second focus group addressing disruption readiness. This phase included preparing AI for out-of-distribution events using synthetic scenarios and digital twins.

A key contribution from the focus groups and interviews is the introduction of three agent design archetypes that clarify the practical applications of AI for supply chain resilience. Diagnostic and Monitoring Agents function as early-warning systems, continuously tracking metrics and identifying slow-moving risks before escalation. Response and Coordination Agents support disruption recovery by retrieving and synthesizing fragmented data across disconnected systems and taking limited actions subject to rigorous human validation at every critical step. Interface and Learning Agents facilitate the dissemination of tacit knowledge and reinforce feedback loops, ensuring that lessons learned from disruptions inform future planning.

The primary deliverable of this thesis is a two-dimensions evaluation framework. The first dimension assesses whether a proposed AI concept credibly enhances supply chain resilience, while the second dimension determines the appropriate level of autonomy. Only concepts that pass the resilience assessment proceed to the autonomy evaluation. The framework serves as more than a technology selection tool; it provides practitioners with a structured method for identifying the specific requirements and conditions under which a control architecture is warranted. By employing a Dimensions Interaction Matrix that maps resilience value against governability across six decision zones, ranging from “Reject” to “Bounded Autonomy,” the framework prescribes the minimum viable automation level and corresponding oversight design. This approach ensures that autonomy does not exceed the organisation’s demonstrated capacity for governance, monitoring, and rollback.

The three archetypes, the two evaluation dimensions, and their integration into a repeatable decision framework constitute an original contribution bridging AI capability research and supply chain resilience theory. The semi-structured interviews effectively identified practitioner and governance prerequisites. The two focus groups were particularly valuable as co-design sessions: the first translated abstract findings into actionable agent portfolios with “always allowed / never allowed” boundaries, while the second stress-tested these concepts against disruption readiness criteria. The cross-domain analogy with Airbus tested whether the eight transferable design pillars for governing AI are anchored in estab-

lished aviation practices, which have undergone decades of improvement in automation best practices. However, this analogy has limitations, as aviation benefits from deterministic architectures and mandatory certification, whereas supply chains are heterogeneous and commercially negotiated. Therefore, these pillars should be regarded as adaptable heuristics rather than prescriptive rules. In this thesis, the concept of “balance” is defined as the explicit calibration of autonomy against the organisation’s demonstrated governance capacity at each decision point. These findings yield three actionable recommendations for practitioners. First, managers are advised to implement a layered control architecture rather than full end-to-end automation, situating AI capabilities where they provide the greatest resilience and maintaining human oversight at critical decision points. Second, investment strategies should emphasize disruption readiness rather than focusing exclusively on operational efficiency, allocating resources to test agent performance in novel scenarios rather than relying solely on historical data. Third, effective deployment requires organizational preparedness that extends beyond technical integration. This includes establishing explicit protocols for human–AI collaboration, defining clear escalation thresholds, and ensuring workforce competencies in both AI literacy and disruption management as prerequisites for responsible adoption.

Contents

1

Introduction

The global trading system continues to experience significant strain, increasingly influenced by both policy and physical risks. Following an election-heavy 2024, political debates are translating into tangible operational challenges and costs for supply chains. The rise of protectionism, the implementation of frameworks such as the EU Carbon Border Adjustment Mechanism (CBAM), and the expansion of carbon markets in China are contributing to greater regulatory complexity in global trade (S&P Global Market Intelligence, 2024). Concurrently, frequent extreme weather events, including disruptions to global storm tracks caused by La Niña, are actively undermining logistics chains (S&P Global Market Intelligence, 2024). Practitioner reports identify climate disruptions, geopolitical tensions, trade policy, and cyber risk as major threats shaping supply chain operations in 2026 (DHL & Analytics, 2025). In response to these challenges, supply chain resilience - defined by Ivanov (2021b) as the ability to maintain a firm's operational continuity despite uncertainties, risks, disruptions and disturbances - has become a central concern for both managers and researchers. In practice, resilience in 2026 is increasingly characterized by the capacity to navigate risks within stringent policy, cost, and weather constraints, rather than by the elimination of risk (Hosseini et al., 2019; OECD, 2025).

A central practical challenge for supply chain managers is determining where and how to intervene. OECD (2025) contends that resilience must be balanced with efficiency and sustainability, necessitating explicit trade-off decisions throughout the supply chain network. Within this context, industry surveys consistently identify artificial intelligence (AI) as a top strategic priority, closely followed by resilience. According to ASCM's criticality index, AI achieves a score of 68.18 out of 100, reflecting leaders' assessments of both its impact and urgency, and indicating a strong near-term influence on supply chains (Association for Supply Chain Management, 2025). Nevertheless, firms adopt a strategic approach, often prioritizing diversification and targeted digital initiatives when the anticipated benefits are clear and immediate. In contrast, comprehensive network redesigns typically demand extended timelines, greater coordination, and significant organizational change (SAP, 2025b; S&P Global Market Intelligence, 2024). Systematic reviews corroborate these findings, demonstrating that AI and machine learning yield the greatest value when problems are well-defined and supported by robust, integrated, and reliable data foundations (Culot et al., 2024; Toorajipour et al., 2021).

Within the broader trajectory of AI adoption, Agentic AI represents a substantial advancement in both capability and autonomy. Agentic AI encompasses systems that coordinate multiple specialized agents via an orchestration layer, enabling autonomous pursuit of objectives, persistent memory retention, and continuous replanning with minimal supervision (Ren et al., 2025; Sapkota et al., 2025). This development marks a transition for AI systems from primarily supporting decision-making to actively executing tasks. However, increased autonomy introduces additional risks. As autonomy expands, the probability of cascading errors and opaque decision-making also rises, particularly when data quality deteriorates or disruptions exceed the system's learned parameters.

Accordingly, this thesis employs a cross-domain analogy by analyzing best practices from aviation, a sector that has benefited from decades of automation development, to extrapolate principles for the safe deployment of Agentic AI in supply chains. Furthermore, it addresses the practical question of when Agentic AI is justified for enhancing resilience and how it should be designed to ensure that autonomy remains aligned with data readiness, operational risk, and organizational capacity for control. From a design and governance perspective, Agentic AI is conceptualized as a layered control problem rather than as an isolated technology. The central argument is that, while Agentic AI holds significant promise for managing disruption, it cannot function effectively in isolation. In the absence of robust data infrastructures and human-in-the-loop safeguards, increased autonomy may introduce additional complexity instead of delivering the targeted, short-term resilience required by organizations.

2

Literature Review

The literature review comprises five main sections. The first section establishes the conceptual foundation of supply chain resilience by defining core terminology such as uncertainty, risk, disturbance, and disruption, and by outlining the temporal dynamics of recovery following disruptive events. This section also situates these concepts within both practitioner and academic perspectives, concluding with an overview of frameworks, tools, and metrics developed to operationalize resilience (Section 2.1). The second section reviews principal supply chain reference frameworks and justifies the adoption of the SCOR model (Section 2.2). The third section analyzes the role of digital technologies, advanced analytics, and artificial intelligence as enablers of supply chain resilience, drawing on empirical evidence regarding their impact on visibility, agility, flexibility, redundancy, and collaboration (Section 2.3). The fourth section provides technical and taxonomic grounding for the AI systems discussed throughout the thesis, including definitions, capability levels, learning paradigms, implementation patterns, and the progression from GenAI to AI agents and Agentic AI systems. The final section identifies research gaps in the design and governance of Agentic AI for supply chain resilience and formulates the research question and sub-questions (Section 2.8).

2.1. Supply Chain Resilience: Concepts, Definitions, and Frameworks

This section establishes the conceptual foundation for the remainder of the thesis. It is structured around a key analytical distinction: the difference between the emergence of adverse conditions and the subsequent supply chain response. The first two subsections address these dimensions sequentially. The initial subsection defines the escalation chain from uncertainty to disruption, introducing the terminology employed throughout the thesis. The following subsection examines the temporal dynamics of recovery after a disruptive event, including the metrics and trajectories through which performance is restored or reconfigured. Building on this conceptual groundwork, the section reviews both practitioner and academic perspectives on resilience and concludes with an overview of the frameworks, tools, and metrics developed to operationalize resilience.

From Uncertainty to Disruption: The Escalation Chain

The definition of the scope of supply chain resilience given by Ivanov (2021b) is to maintain a firm's operational continuity despite uncertainties, risks, disruptions and disturbances. According to Ivanov and Sokolov (2010) in the analysis of uncertainty and risks four aspects can be encountered, as shown in Figure 2.1.

Uncertainty is a property of the system environment: incomplete or imperfect knowledge about the system, its context or how conditions will develop. It exists independently of particular decisions and can be broadened or narrowed through additional information, modeling or learning. Some uncertainty is reducible (data, experiments or learning can shrink it); some is irreducible (deep, unknown-unknowns). Example: ambiguous future demand patterns or uncertain geopolitical conditions.

Risk arises from uncertainty and expresses it in causal, decision-relevant terms: identifiable possible events or outcomes together with their estimated likelihoods and consequences. Risks can be identified, analyzed, prioritized and managed (for example through mitigation, control, transfer or acceptance measures). Example: the risk of demand fluctuation that stems from broader market uncertainty.

Disturbance is the realized perturbation that follows from one or more risks - a concrete, observed impact on operations, processes or flows. Disturbances may be intentional (for example theft or sabotage) or unintentional (for example sudden demand swings). Whether a disturbance produces further harm depends on system properties such as robustness and adaptability: a disturbance may be absorbed without causing lasting effects, or it may propagate through the network.

Disruption is a disturbance that produces a significant, adverse deviation from planned operations or network functioning - a breakdown of normal continuity that typically requires adaptation or recovery actions. Disruptions can be transient and local or structural and severe (for example prolonged supplier unavailability). Example: a supplier shutdown that halts production and triggers downstream shortages.

These concepts form a causal chain in practice: uncertainty $\rightarrow$ risk $\rightarrow$ disturbance $\rightarrow$ disruption. Not every risk materialises as a disturbance, and not every disturbance escalates to a disruption; system properties and managerial responses determine how far an event propagates.

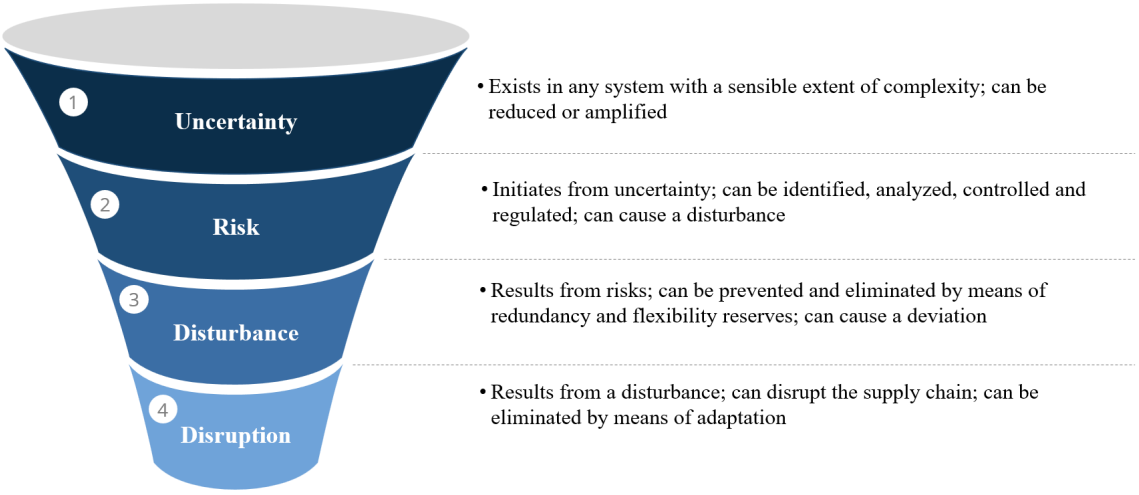


Figure 2.1: Definition of the scope of supply chain resilience given by Ivanov and Sokolov (2010).

Figure 2.1 synthesizes the escalation logic from uncertainty to disruption. This representation clarifies the origin and intensification of adverse events, but it does not yet capture how the supply chain behaves once disruption has occurred. That temporal dimension is addressed in the following section through the concept of recovery.

The concepts above define the escalation chain through which adverse conditions emerge and intensify, from ambient uncertainty to operational disruption. However, identifying how a disruption arises is analytically distinct from understanding what happens after it occurs. Once a disruption materialises, the focus shifts from event classification to system response over time, namely the mechanisms and speed through which supply chain performance is restored, adapted, or reconfigured.

From Disruption to Recovery: Restoring Supply Chain Performance

While the previous section distinguished the escalation chain from uncertainty to disruption, this section examines the temporal dimension that follows disruption: recovery. Recovery is not simply another disturbance-related concept; it represents the process through which the supply chain restores, stabilizes, or redefines its performance after a disruptive event. Hohenstein et al. (2015) formalize this temporal logic by identifying four distinct phases of supply chain resilience: readiness (anticipating and preparing for risks), response (reacting and adapting during a disruption), recovery (restoring operations to at least pre-disruption levels), and growth (learning and improving beyond the prior state). Recovery time, defined as the duration from the onset of a disruption to the point at which a target performance level is regained, has become a widely used quantitative proxy for resilience. Simchi-Levi et al. (2014) operationalize this concept through two complementary metrics: Time-to-Recover (TTR), the time required for a supply chain node to be restored to full functionality after a disruption, and Time-to-Survive (TTS), the maximum duration the supply chain can continue matching supply with demand while a node is inoperative. A supply chain is considered robust at a given node when its TTS exceeds the node's TTR (Simchi-Levi et al., 2018). Recovery is shaped by both pre-disruption design choices, such as redundancy, flexibility, and inventory buffers, and by post-disruption managerial actions, including re-routing, supplier switching, and demand management (Chowdhury & Quaddus, 2017a).

The Temporal Dynamics of the Resilience Cycle

The temporal dynamics of the resilience cycle are demonstrated by the disruption profile introduced by Sheffi and Rice (2005), as shown in Figure 2.2. This profile depicts supply chain performance over time across eight sequential phases: preparation, disruptive event, first response, initial impact, full impact, recovery preparations, recovery, and long-term impact. The shaded area between the baseline and the actual performance curve represents cumulative performance loss. Two characteristics of this area are fundamental to resilience assessment: its depth, which indicates disruption severity, and its length, which denotes recovery time.

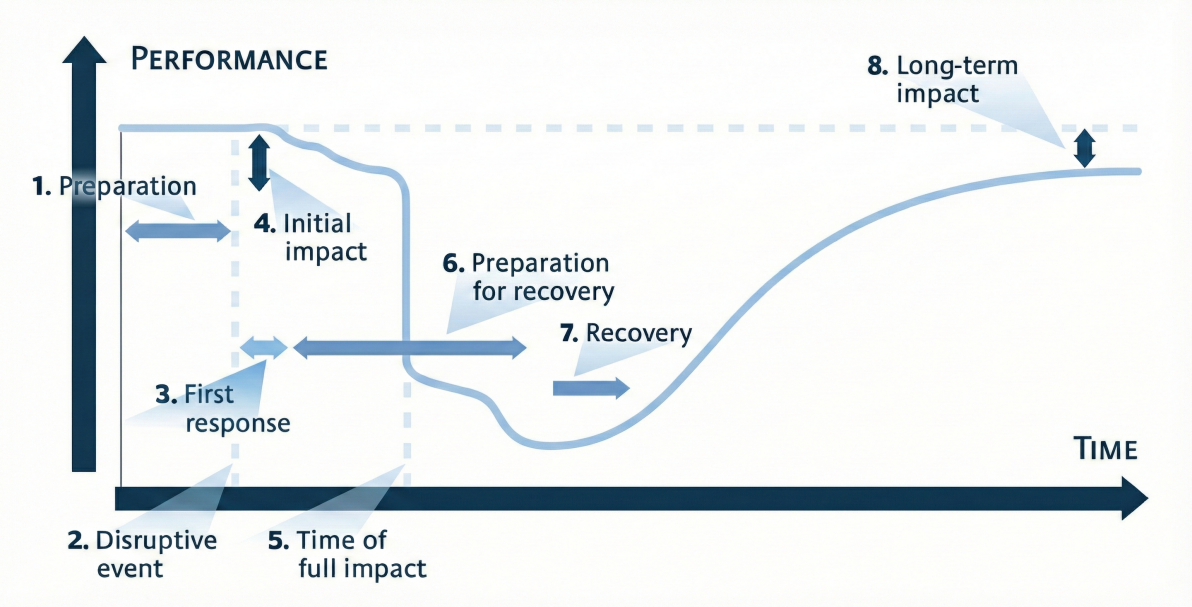


Figure 2.2: The disruption profile depicting supply chain performance over time through eight phases Sheffi and Rice (2005).

This shaded area corresponds directly to what Bruneau et al. (2003) formalized as the resilience triangle, a geometric representation of cumulative performance loss originally developed for seismic infrastructure resilience and later adapted to supply chains by Falasca et al. (2008). The primary objective of supply chain resilience strategies is to minimize the size of this triangle, either by limiting the depth of the performance decline, by shortening the recovery period, or by achieving both outcomes simultaneously (Falasca et al., 2008). As noted by Hohenstein et al. (2015) and Hendricks and Singhal (2005), the supply chain resilience literature indicates that post-disruption recovery may follow various trajectories, including recovery to an improved baseline or persistent performance loss. Instead of merely restoring the previous baseline, recovery may exhibit a ”bounce forward” trajectory, which aligns with the definition of resilience as a system’s capacity to not only return to its original state but also to adapt, grow, and transition to a more desirable state following disturbance (Hohenstein et al., 2015). Effective recovery also requires maintaining continuous viability, which means that interconnected supply networks must dynamically adapt their behaviors and structures to persist in a changing environment (Ivanov & Dolgui, 2020). Alternatively, the recovery trajectory may be complicated by after-shock risks (Ivanov, 2021a). During an evolving crisis, resuming operations prematurely can result in renewed disruptions. For instance, hastily restarting production without sufficient risk mitigation may trigger new disruption clusters and subsequent operational shutdowns (T. Fan et al., 2023). Additionally, supply chains are exposed to the risk of ”disruption tails,” which are postponed effects or residual impacts from the initial disruption period, such as backlogs and delayed orders (Ivanov, 2021a). If management policies are not appropriately adapted, these disruption tails can significantly destabilise inventory and production dynamics during the post-disruption recovery phase (Ivanov, 2021a).

Practitioner Perspectives

Recent discourse among supply chain leaders has reframed resilience as a proactive and adaptive capability, developed through pre-disruption investments in flexibility, visibility, and collaboration. This approach enables organizations to maintain operations and potentially enhance performance during

disruptions, aligning with research that defines resilience as the capacity to survive, adapt, and grow, and that advocates for both proactive and reactive capabilities (Chowdhury & Quaddus, 2017b; Pettit et al., 2010). Analysts observe a shift from precise, single-point planning to probabilistic and flexible planning, with the aspirational objective of achieving 'antifragility' - networks that not only withstand volatility but also benefit from it (Payne, 2025). Professional bodies similarly converge on a definition of resilience as the capability to anticipate, adapt, and recover from disruptions in ways that protect or enhance competitive position, emphasizing preparedness before disruptions, agility during them, and recovery afterward (ASCM, n.d.). Technology vendors reinforce these points but place particular emphasis on end-to-end visibility and scenario planning as practical enablers, contending that connected planning, flexible resource reallocation, and rapid re-forecasting are essential for resilient responses from sourcing through final delivery (SAP, n.d.).

Executive guidance from technology firms extends this perspective by emphasizing the importance of digitization and workflow automation. Resilience is characterized as the ability to anticipate, adapt, and recover from a wide range of shocks, including pandemics and geopolitical events. Digital tools are identified as providing the situational awareness and coordinated execution necessary to translate strategic intent into operational action (IBM, n.d.-b). This aligns with the views of senior executives, who maintain that resilience should be balanced with efficiency, rather than pursued as an isolated objective, in light of ongoing geopolitical instability, increased transit times, and recurring delivery risks (Deloitte Insights, 2024). Strategy research further refines this balance by advocating for a "cost of resilience" approach, in which companies redesign manufacturing and sourcing networks to remain flexible during disruptions without compromising margins, explicitly weighing agility against cost competitiveness (Boston Consulting Group, 2025). Policy shocks, such as tariffs, add complexity to location decisions. Cost models identify clear tipping points at which reshoring or nearshoring becomes economically viable, positioning resilience as both a capital allocation and operational decision (Deloitte Insights, 2025). Concurrently, enterprise risk perspectives frame resilience as a strategic lever that should be integrated into commercial and boardroom discussions, given the increasing interdependence with suppliers and third parties and the corresponding need for transparency and cross-functional collaboration (Deloitte UK, 2022).

Academic Perspectives

Academic literature has both paralleled and enriched the practitioner debate by tracing the evolution of supply chain resilience from early conceptualizations to advanced modeling and digital transformation. A significant contribution by Christopher and Peck (2004) defined resilience as the capacity of a supply chain to maintain continuity during turbulence, emphasizing visibility (end-to-end awareness of inventory, demand, and material flows), velocity, flexibility, and collaboration. This work marked a shift from viewing risk management as the prevention of known threats to understanding resilience as an adaptive and systemic property. Concurrently, Sheffi and Rice (2005) used field evidence to argue that resilience requires both redundancy and flexibility, and emphasized that resilient enterprises integrate risk considerations into daily decision-making rather than confining them to specialized risk functions.

Similarly, Ponomarov and Holcomb (2009) synthesized definitions from psychology, ecology, and engineering to define resilience as the adaptive capability of a supply chain to prepare for, respond to, and recover from disruptions. This perspective emphasized that resilience is not simply a return to a previous state, but an emergent capacity that can enhance performance. Tukamuhabwa et al. (2015) conducted a comprehensive review of ninety-one contributions, highlighting the absence of a unified

definition. The review identified flexibility, redundancy, agility, and collaboration as dominant strategies, and recommended Complex Adaptive Systems (CAS) theory as a suitable theoretical framework given the non-linear and interconnected characteristics of supply chains.

Frameworks, Tools, and Metrics for Assessing Supply Chain Resilience

Kamalahmadi and Parast (2016) reinforced this multidimensional understanding in their review of one hundred studies. They distinguished resilience from risk management, stressing that traditional frameworks fail to address unpredictable disruptions. They defined resilience as the ability of organizations to survive, adapt, and grow despite uncertainty, highlighting the tension between efficiency-oriented practices and the need for robust capabilities. Pettit et al. (2010) added a conceptual framework identifying seven vulnerabilities, such as turbulence and resource limitations, alongside fourteen resilience capabilities, including flexibility, visibility, and collaboration. Their central argument was that resilience arises from balancing vulnerabilities and capabilities rather than maximizing one at the expense of the other.

The same authors later developed the SCRAM™ tool to operationalize their framework. Tested with seven multinational firms and fourteen real-world disruptions, it documented more than 300 linkages between vulnerabilities and capabilities, demonstrating a positive link between resilience strength and supply chain performance (Pettit et al., 2010). Their retrospective article a few years later emphasized that resilience research must evolve from a view of bouncing back to one of bouncing forward, using crises as opportunities for transformation (Pettit et al., 2019). A similar perspective was offered by Wieland and Durach (2021), who contrasted engineering resilience, focused on stability and recovery, with social-ecological resilience, which emphasizes adaptability and transformation. They argued that supply chain research should integrate both perspectives to better address unexpected events, such as the COVID-19 pandemic.

Beyond these conceptual frameworks and their operationalization, scholars have also sought to measure and model resilience empirically. Jain et al. (2017) proposed a hierarchical model identifying thirteen enablers of resilient supply chain practice, including adaptability, trust, and information sharing. Using interpretive structural modeling and structural equation modeling with data from Indian firms, they introduced a resilience index that allows benchmarking across supply chains. Behzadi et al. (2020) approached measurement through financial performance, critiquing traditional metrics such as Time to Recovery and Recovery Level. They proposed the Net Present Value of Lost Profit (NPV-LP) as a more holistic measure that integrates both time and financial impact, demonstrating that the choice of metric can decisively shape managerial decisions.

The complexity of global sourcing has also been recognized as a source of both risk and resilience potential. Gunasekaran et al. (2015) argued that while global sourcing lowers costs and provides access to innovation, it increases vulnerability through longer lead times and geopolitical risks. However, a globally diversified sourcing network can also strengthen resilience, as evidence from the COVID-19 pandemic showed that firms sourcing across multiple countries were able to shift procurement volumes and activate alternative suppliers in different regions, leveraging the breadth of their global sourcing base as a buffer against localized disruptions (Niu et al., 2025). More recently, attention has shifted to the role of digitalization; Zhao et al. (2023) demonstrated empirically that digitalization enhances supply chain resilience by strengthening visibility, coordination, and responsiveness, with resilience serving as a mediating mechanism between digitalization and performance.

2.2. Supply Chain Reference Framework Selection and SCOR Model

According to Lambert and Cooper (2000) and Ntabe et al. (2015) the main families used to represent and manage supply chains in research and practice are: governance/process frameworks such as the *Global Supply Chain Forum (GSCF)*; enterprise process classifications (PCF-type taxonomies); collaboration protocols such as *CPFR*; strategy-to-performance architectures such as the *Balanced Scorecard (BSC)*; value/design-chain references such as *VCOR/DCOR*; and operations reference models such as *SCOR*.

In this literature, *GSCF* formalizes eight cross-functional, cross-firm processes and is strong on relationship governance and process ownership, but it is less prescriptive at activity level and does not foreground a standardized, supply chain-specific metric canon for process-by-process benchmarking (Lambert & Cooper, 2000; Lambert & Enz, 2017). PCF-style taxonomies provide a robust common language and hierarchical breakdown that aid alignment and generic benchmarking, yet they operate primarily as classification scaffolds rather than domain diagnostics tightly coupled to Plan–Source–Make–Deliver–Return metrics for intervention evaluation (Svensson & Hvolby, 2012). *CPFR* is a focused collaboration method around planning/forecasting/replenishment; empirical and simulation studies report operational and financial gains and advantages over alternatives such as VMI, confirming its value for demand–supply alignment while remaining deliberately narrow and not an end-to-end process-plus-metrics reference across Source–Make–Deliver–Return (Hill et al., 2018; Sari, 2008). *BSC* remains influential for strategy-to-performance alignment and KPI cascades, but in operations studies it is typically paired with a domain process model when activity-level diagnosis and benchmarkable trade-offs are required, because BSC itself is not a supply chain process hierarchy (Balaji et al., 2021; Tawse & Tabesh, 2023). *VCOR/DCOR* widen scope toward customer/product value and upstream design/R&D and appear mainly in simulation or reference-modeling contexts - often starting from SCOR templates - indicating conceptual breadth but a thinner, practice-tested operational metric canon for routine diagnostics (Persson, 2011; Svensson & Hvolby, 2012).

By contrast, The Supply Chain Operations Reference (SCOR) model, which was introduced in the 1990s, provides a structured framework to evaluate the effectiveness and efficiency of supply chains in relation to sales and operations planning (S&OP). Given the complexity of supply chain management and the challenges associated with S&OP implementation, the model serves to standardize processes and monitor in a systematic and measurable way the performance outcomes, also across diverse industries (G. Stewart, 1997; White, 2025). *SCOR* is consistently documented as an end-to-end framework with explicit Level-2/Level-3 decomposition (Plan–Source–Make–Deliver–Return–Enable) *and* a standard performance system (attributes → metrics) that enables benchmarking and multi-criteria prioritization; reviews and applications show SCOR used as a diagnostic/redesign backbone (including environmental "GreenSCOR"), paired with decision methods to prioritize process-level improvements and with modern analytics/ML to *predict* performance from SCOR metrics; SCOR also underpins configuration/simulation tooling (Huang et al., 2005; Lima-Junior & Carpinetti, 2019; Ntabe et al., 2015; Palma-Mendoza, 2014; Persson, 2011).

In line with the needs stated in Chapter 3, a single end-to-end map with clear, decision-level steps - a standard and benchmarkable metric set to compare heterogeneous KPIs, and a well-validated base that is also extensible to resilience and sustainability - the frameworks play complementary roles (GSCF for governance; PCF for shared language; CPFR for collaboration in planning; BSC for strategy discipline; VCOR/DCOR for value/design scope). However, *SCOR* is the only one that brings these pieces

together: comparable process definitions across *Plan–Source–Make–Deliver–Return–Enable*, a mature, field-tested metric library, and broad applied evidence. We therefore adopt SCOR as the organizing ”reference language” and framework for the entire study, providing a single, practical structure that keeps technical and operational work aligned. It will shape the interview protocols so evidences from AI specialists and supply chain practitioners are directly comparable, organize the AI concept briefs by process, and map resilience areas (visibility, agility, flexibility, redundancy, and collaboration) onto specific processes. This ensures every proposed change is anchored to concrete decision points and Level-2 process elements rather than abstract capability descriptions, and enables assessment of where AI yields measurable improvements e.g., forecasting in *Plan*, supplier-risk sensing in *Source*, predictive maintenance in *Make*, control-tower/dynamic routing in *Deliver*, closed-loop diagnostics in *Return*, and cross-cutting data governance/cybersecurity/talent in *Enable* (APICS/ASCM, 2017; Lima-Junior & Carpinetti, 2019; Ntabe et al., 2015; Palma-Mendoza, 2014; Zhou et al., 2011).

SCOR is organized into six core process categories: *Plan*, *Source*, *Make*, *Deliver*, *Return*, and *Enable* (APICS/ASCM, 2017). *Plan* aligns resources, requirements, and communication with business goals. *Source* covers the procurement of goods and services to meet demand. *Make* transforms inputs into finished products through production activities. *Deliver* manages orders, transportation, and distribution to fulfil demand. *Return* handles reverse logistics and product returns. Finally, *Enable* supports execution with business rules, performance monitoring, data, contracts, compliance, and risk management (White, 2025).

The model is hierarchical: Level 1 defines the scope, content, and performance targets of the supply chain; Level 2 specifies process categories and sets the operations strategy; Level 3 details individual process elements, metrics, and capabilities; and Level 4 connects to improvement tools and activities such as lean, six sigma, benchmarking, and best practices (APICS/ASCM, 2017). Several practice studies support the actual functioning of the SCOR model’s structure. Zhou et al. (2011) found that the *Plan* process has a direct positive effect on *Source*, *Make*, and *Deliver*. This demonstrates how good planning supports later steps in the chain. Moreover, Lockamy and McCormack (2004) demonstrated that planning and collaboration based on SCOR enhance performance, particularly when supported by clear process measures, system integration, and information technology.

SCOR also provides a base for measuring performance; in fact, it links key attributes like reliability, speed, flexibility, cost, and asset use to standard metrics, and this simplifies the comparison and prioritization across companies. To bridge strategy and operations, these decision tools use Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) together with the SCOR metric structure to give each strategic goal a weight and to make different kinds of measures comparable, that provides a practical way to turn strategy into clear operational trade-offs (Kocaoğlu et al., 2013). Moreover, SCOR has been applied and extended, in particular the growing importance of *Return* and the environmental dimension (often termed GreenSCOR) (Ntabe et al., 2015). Existing studies demonstrate how organizations use SCOR and the competitive effects they report, showing its value as a common framework for process redesign and performance benchmarking across Levels 1–3 (Delipinar & Kocaoglu, 2016).

Seen in this light, SCOR provides a practical connection to current resilience and digital transformation priorities. As organizations move from simple linear supply chains to networked, data-rich models, industry guidance still treats SCOR as the core process map for modernising capabilities and metrics (Deloitte & ASCM, 2020). However, the capabilities embedded in this framework - visibility, agility,

adaptive planning - increasingly depend on digital technologies and advanced analytics. Section 2.5 examines this digital transformation and its implications for resilience.

2.3. Digital, Analytical and AI Enablers of Supply Chain Resilience

Recent literature shows that digital technologies, advanced analytics and artificial intelligence are becoming important enablers of supply chain resilience. Studies discuss digital indicators and stress-testing tools, optimization and simulation models, blockchain-based visibility, and the use of AI and big data analytics to strengthen capabilities such as visibility, agility, flexibility, redundancy and collaboration. This section brings these strands together and provides the technological and analytical context for the later focus on Agentic AI.

Digital and AI-Driven Enablers for Supply Chain Resilience

Recent research demonstrates that digital technologies, big data analytics, and artificial intelligence can enhance supply chain resilience by improving visibility, agility, flexibility, redundancy, and collaboration (Naz et al., 2022; Zamani et al., 2023). Emerging technologies further expand these capabilities. Meafa et al. (2024) investigated Metaverse applications and GenAI (GenAI), finding that immersive simulations and AI-driven scenario generation facilitate collaboration and proactive decision-making, but also introduce concerns related to security, ethics, and governance. Similarly, Boone et al. (2025) differentiates GenAI from traditional prediction- and optimization-focused AI, stating that ”unlike traditional AI, which relies on historical data for prediction and optimization, GenAI can generate novel solutions and simulate alternative scenarios in real time.” However, they warn of risks such as hallucination, misinformation, security, and privacy, and stress the importance of governance and human-in-the-loop validation. These advancements require epistemic caution. As Bender and Koller (2020) observe, ”the language modeling task, because it only uses form as training data, cannot in principle lead to learning of meaning.” National Institute of Standards and Technology (2024b) further clarifies that large language models ”generate outputs that approximate the statistical distribution of their training data” by ”predicting the next token or word.”

In addition to GenAI, established AI and analytics approaches offer proven pathways to resilience. Belhadi et al. (2022) developed a decision-making framework that integrates AI techniques with multi-criteria decision making, demonstrating how specific technologies align with resilience levers and highlighting the value of hybrid approaches when data are noisy or incomplete. Systematic reviews support these conclusions. Zamani et al. (2023) finds that AI and big data analytics consistently improve visibility and agility, but also identifies ongoing challenges related to ethics, explainability, and the integration of human judgment with algorithmic recommendations under time constraints. Likewise, Naz et al. (2022) contends that post-COVID-19 resilience relies on data-driven supply chains capable of responding to rapidly changing conditions, emphasizing that without reliable data quality, information sharing, and analytics capability, AI solutions cannot achieve their intended outcomes.

Recent research delineates the domains in which digital and AI-driven technologies provide the most immediate value. Beta et al. (2025), through a systematic review of 25 studies, identifies predictive analytics for demand and supply forecasting, AI-powered risk management systems that assess threats across financial, operational, and geopolitical dimensions, and automated real-time monitoring for anomaly detection as the AI-driven technologies most consistently strengthening supply chain resilience. The operational impact of these technologies is considerable, with reported improvements of

35% in inventory levels, 15% in logistics costs, and 65% in service levels. These results are empirically supported by Belhadi et al. (2024), whose survey of 279 manufacturing firms demonstrates that AI-driven information processing capabilities significantly enhance supply chain performance. Although AI directly influences performance, the study indicates that the effect mediated through supply chain resilience is even more pronounced, primarily due to AI's capacity to enhance adaptive capabilities and foster supply chain collaboration. Transitioning from aggregate evidence to specific applications, Riad et al. (2024) presents empirical data on the operational impact of AI in enhancing supply chain resilience. In an electronics manufacturing case study, its integration reduced stockouts from 9% to 2% and shortened average delivery times from 8.5 to 4.2 days. The proposed framework further indicates that GenAI improves resilience by generating detailed what-if simulations to evaluate mitigation strategies for potential disruptions prior to their occurrence. At the strategic level, D. Singh et al. (2025) utilize the dynamic capabilities perspective to illustrate how particular algorithms facilitate resilience. For example, deep learning is used to analyze Point-of-Sale data to identify alternative suppliers, while metaheuristic algorithms dynamically reconfigure transportation routes during logistical blockages. This approach is consistent with the Industry 5.0 perspective advanced by Ahmed et al. (2023), who quantitatively identify real-time tracking via the Internet of Things (IoT) as the most critical AI-based imperative. This application is argued to enhance supply chain resilience by providing inventory visibility for immediate route optimization.

However, significant challenges persist. Reviews in the *International Journal of Supply Chain Management* report that fragmented data and siloed systems frequently hinder the scaling of AI pilot projects, emphasizing that comprehensive data capture is necessary for substantial resilience improvements (Kumawat, 2024). Governance and security are equally critical. The *International Journal of Scientific and Management Research* underscores that AI-driven resilience depends on reliable data pipelines (Chukwu et al., 2024). Collectively, the evidence indicates that AI and big data analytics enhance resilience when organizations prioritize data quality, align methods with resilience objectives, and integrate outputs into decision-making processes. Forecasting, optimization, and anomaly detection represent the most immediate applications, while human - AI collaboration, governance, and partner coordination are essential for sustained adaptability (Belhadi et al., 2022; Riad et al., 2024).

2.4. Defining Artificial Intelligence

Previous sections have defined supply chain resilience and identified the locations of resilience-related decisions within the SCOR process structure. To determine which artificial intelligence (AI) approaches are suitable at these decision points, and to differentiate between conventional prediction and optimization, GenAI, Agents, and Agentic AI systems, a precise definition of AI is necessary. This section presents the working definition of AI adopted in this thesis.

What is AI?

In the literature, there is no single, universally accepted definition of artificial intelligence (AI). The term artificial intelligence was first coined in the 1956 Dartmouth proposal, a two–month, ten–person summer research project submitted by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. In that proposal, AI was framed as the problem of making a machine "behave in ways that would be called intelligent if a human were so behaving" and of finding how to make machines use language, form abstractions and concepts, solve problems then reserved for humans, and improve

themselves (McCarthy et al., 2006). In a later and more concise formulation, McCarthy (2007) defined AI as the science and engineering of making intelligent machines, especially computer programs, capable of achieving goals in the world. In this view, intelligence is understood as the computational ability to reach goals and can appear in different forms and degrees in humans, animals, and machines; moreover, AI is not restricted to biologically inspired methods, but may exploit procedures very different from human cognition.

Russell et al. (1995) further systematize the concept of AI by distinguishing four complementary viewpoints: acting humanly, as in the Turing Test, where a machine's behavior is indistinguishable from that of a human interlocutor; thinking humanly, where AI models the cognitive processes studied in psychology and cognitive science; thinking rationally, where systems follow formal rules of logic to derive correct conclusions; and acting rationally, where AI is conceived as a rational agent that perceives its environment and selects actions that maximize its chances of success.

2.5. Taxonomies of AI

Capability Levels, System Boundaries, and Learning Paradigms

A widely used taxonomy classifies artificial intelligence according to its capabilities relative to human intelligence. Artificial Narrow Intelligence (ANI) describes systems that excel at specific tasks but cannot function beyond their designated domains. Examples include demand forecasting models, route optimization algorithms, quality inspection systems, and predictive maintenance tools. Broad AI characterizes the current enterprise landscape, where multiple narrow systems are integrated into business processes to address a broader range of related tasks using domain-specific knowledge and data, yet still lack general intelligence (Feng et al., 2024; C. S. Singh et al., 2019). Artificial General Intelligence (AGI) refers to hypothetical human-like intelligence capable of reasoning and learning across diverse domains, whereas Artificial Superintelligence (ASI) would surpass human capabilities in all respects (Strelkova, 2017).

Contemporary artificial intelligence systems are advancing beyond standard generative large language models (LLMs) and are increasingly described as "large reasoning models" (LRMs). Although these systems are based on LLM architectures, frontier models such as OpenAI's o1 are distinguished by their use of chain-of-thought processing as a core mechanism, enabling them to engage in reasoning prior to response generation. These models exemplify advanced forms of Narrow or Broad AI, surpassing traditional generative methods while remaining distinct from artificial general intelligence (AGI). Large-scale evaluations using the AGI-Benchmark 1.0 indicate that o1 achieves human-level or superior performance on a range of complex tasks across diverse domains, including competitive programming, mathematical derivation, medical diagnosis, and other challenges involving complex reasoning, planning, and diagnostics. Nevertheless, these assessments also highlight ongoing limitations in abstract reasoning, adaptability to dynamic environments, and the management of domain-specific edge cases. Collectively, these findings suggest that current frontier systems are positioned between advanced Narrow/Broad AI and true AGI, rather than having achieved general intelligence (Zhong et al., 2024).

At the implementation level, AI systems utilize three primary learning paradigms (Russell & Norvig, 2021). Supervised learning involves training on labeled input-output pairs to predict outcomes for new data; in supply chains, this approach supports demand forecasting, lead-time prediction, and

quality classification. Unsupervised learning identifies patterns in unlabeled data without predefined categories, enabling supplier risk clustering, detection of shipment delay patterns, and anomaly identification within logistics networks. Reinforcement learning operates through interaction with the environment, where the system takes actions, observes outcomes, and receives rewards or penalties to refine its policy over time. This paradigm facilitates dynamic inventory allocation, adaptive routing during disruptions, and real-time sourcing decisions that balance multiple objectives.

Collectively, these learning paradigms explain how AI systems process data and feedback to adapt their behavior. The selection of a particular paradigm depends on factors such as data availability (labeled versus unlabeled), task structure (prediction, optimization, or sequential decision-making), and operational context. These considerations directly influence which AI approaches are suitable for specific supply chain resilience capabilities.

Implementation Patterns: Foundation Models, Customization, and RAG

From the point of view of organizations, the most visible advances in AI in recent years come from foundation models: large-scale, general-purpose models trained on very large and often multimodal datasets (for example, text, images, clinical records, and scientific literature). These models learn broad patterns and representations that can be adapted to many different tasks, and in their most advanced forms they underpin many GenAI applications and emerging Agentic AI systems (Medetalibeyoglu et al., 2024). They sit on top of the technical layers described above, but expose higher-level capabilities that can be reused across domains.

In practice, it is rarely convenient or economically viable for most companies to design entirely new models and algorithms from scratch at these lower layers of the stack. Developing bespoke learning algorithms requires highly specialized expertise, substantial research and development investment, and long development cycles, which are often misaligned with business timeframes and resource constraints. Instead, organizations typically work at an intermediate level of customization. According to Diaferia et al. (2022), the customization of AI systems is a multi-layered concept, in which model customization is a separate layer from data, algorithms, and infrastructure. Model customization means creating new AI models by repeatedly training, testing, and comparing several algorithms offered by a given AI product. In this approach, organizations do not design new algorithms from scratch. Rather, they use standard packages, frameworks, and libraries that provide state-of-the-art algorithms, and focus on applying them to their own data. The main effort lies in choosing, training, and tuning these existing algorithms to find the model that best fits the specific business problem. This is different from algorithm customization, where the internal structure or code of the algorithm itself is changed. With model customization, the algorithm is treated as fixed, and the work is done at the level of selection and parameter tuning, either by data scientists or through automated tools. As a result, this type of customization sits at an intermediate level: it goes beyond simply using a fully pre-trained product, but it is still less complex than building new algorithms or infrastructure from the ground up (Diaferia et al., 2022).

Building on this intermediate level of customization, many companies are now making AI more suitable for enterprise use by combining general-purpose language models with retrieval-augmented generation (RAG) architectures. RAG connects a generative model to an external knowledge store (for example, a vector database over internal documents), so that at query time the model conditions its answers on relevant retrieved passages instead of relying only on what is stored in its parameters (Lewis, Perez,

Piktus, Petroni, Karpukhin, Goyal, Küttler, Lewis, Yih, Rocktäschel, et al., 2020; A. Xu et al., 2024). In this way, organizations keep using powerful foundation models as they are, while adapting them to their own context mainly through the curation of the retrieval pipeline and the underlying enterprise data, rather than by training new models from scratch.

A concrete example is provided by Z. Xu et al. (2024), who introduce a customer-service question answering system that combines RAG with a knowledge graph built from issue-tracking tickets. Instead of indexing tickets only as independent text chunks, their approach preserves ticket structure (for example, separating key fields such as descriptions and reproduction steps) and adds links between tickets using both explicit relations in the tracking system and semantic similarity. At query time, the system retrieves relevant tickets and related nodes from the graph and uses this retrieved evidence to support answer generation, so that responses are grounded in the most pertinent parts of the historical record.

Recent surveys show that RAG has become a central pattern for knowledge-intensive and domain-specific applications, because it allows large language models to access up-to-date, domain-specific information and significantly improves factual accuracy and robustness in real-world use (W. Fan et al., 2024; Gao et al., 2023). Enterprises can continuously update their knowledge bases and document indexes without retraining the underlying model, which is especially valuable in fast-changing environments such as finance, healthcare, or legal services. In addition, empirical studies indicate that RAG pipelines help reduce hallucinations and make outputs easier to verify, since responses can be grounded in retrieved evidence from trusted internal sources (W. Zhang & Zhang, 2025). From the perspective of the earlier discussion, RAG can therefore be seen as a practical enterprise mechanism that operates above the algorithmic layer: companies select a base model, configure a retrieval and indexing layer over their proprietary data, and then customize the overall system mainly through data selection, retrieval design, and model prompts, rather than through deep algorithmic or architectural changes.

2.6. From GenAI to Agentic AI

A complementary, system-oriented taxonomy classifies modern AI according to the level of autonomy, planning capability, and tool integration it can exhibit in real workflows. This perspective distinguishes between: (i) GenAI systems that provide single-step responses; (ii) AI agents that autonomously execute bounded tasks through iterative planning and tool use; and (iii) Agentic AI systems that coordinate multiple agents with memory, orchestration, and sustained goal pursuit. The progression reflects a shift from passive content generation to active, multi-step, goal-directed behavior.

GenAI Systems

Generative artificial intelligence (GenAI) refers to models that learn the probability distribution of data and the relations between them to generate new text, images, audio, or code. At the system level, GenAI typically operates through a straightforward interaction pattern: prompt, model inference, and output. The system does not autonomously reformulate tasks, plan, or revise its reasoning beyond the prediction step. Instead, a human specifies the request and validates the outcome. However, newer large reasoning models (LRMs) like OpenAI's o1 represent a significant departure from this pattern. These systems integrate chain-of-thought processing as a fundamental mechanism, enabling them to revise and refine their reasoning across multiple internal steps before generating a response—effectively "thinking through" problems rather than producing immediate outputs (Zhong et al., 2024).

Within this thesis, 'single-step' denotes the absence of autonomous multi-step planning and tool-executing goal pursuit by the system; while multi-turn chat or iterative drafting may occur, these processes remain user-driven rather than agent-driven.

In enterprise applications, GenAI is frequently integrated with retrieval-augmented generation (RAG), a method in which task-relevant information is retrieved from external knowledge bases and incorporated into the model at inference time (Harshvardhan et al., 2020; Lewis, Perez, Piktus, Petroni, Karpukhin, Goyal, Küttler, Lewis, Yih, et al., 2020; Xiao et al., 2024). RAG significantly enhances factual grounding and domain relevance, particularly in documentation-intensive domains such as legal, finance, or customer support (W. Fan et al., 2024; Gao et al., 2023). Empirical studies consistently demonstrate performance improvements in summarization, drafting, translation, and coding assistance (Brynjolfsson et al., 2025; Noy & Zhang, 2023). Field experiments further confirm productivity gains, with the most substantial improvements observed among less experienced workers performing well-defined, text-heavy tasks (Brynjolfsson et al., 2025).

However, despite these advances, GenAI systems may still generate plausible yet incorrect or weakly supported outputs, necessitating human oversight and provenance-aware verification (Barman et al., 2024; Farquhar et al., 2024; Omar et al., 2025; Xiao et al., 2024). Managerial evidence indicates that models alone provide limited value without accompanying sociotechnical efforts such as process redesign, assignment of decision rights, and implementation of measurement systems (Dwivedi et al., 2021; McKinsey & Company, 2025). Consequently, organizations seeking to industrialize GenAI invest in data quality, governance frameworks (including NIST's GenAI Profile and the EU AI Act), and workforce skills to enable responsible scaling (Deloitte AI Institute, 2025; European Commission, DG CONNECT, 2025; National Institute of Standards and Technology, 2024c). The autonomy of these systems remains restricted to single-turn generation rather than multi-step execution, and successful deployment depends on clear task design, human-in-the-loop controls, and integration with existing organizational systems (Akter et al., 2021; Maslej et al., 2025).

AI Agents: Autonomous Execution of Bounded Tasks

AI agents build upon generative models by embedding them within iterative decision-making loops. These systems function through repeated cycles of planning, acting, observing, and revising. Typically, a planner such as a large language model (LLM) decomposes a problem into discrete steps, while a tool interface executes operations using application programming interfaces (APIs) or other external functions. Short-term memory retains intermediate results, allowing the agent to reason over sequences of actions rather than isolated prompts (Piccialli et al., 2025; Wang et al., 2024).

Agents demonstrate autonomy within clearly defined task boundaries. They determine which tools to invoke, when to retrieve external information, and how to adjust subsequent steps based on observed outcomes. Common applications include structured data extraction, document search, code execution, workflow orchestration, and basic analytical tasks. While these systems enhance efficiency and consistency, they continue to require human supervision, restricted permissions, and explicit task definitions (Yu et al., 2025). Consequently, their autonomy remains local and specific to the assigned task.

Agentic AI Systems: Orchestration, Memory, and Multi-agent Coordination

Agentic AI constitutes an advancement in both autonomy and system-level complexity. Instead of depending on a single agent, agentic systems coordinate multiple specialized agents, such as planners,

retrievers, evaluators, and executors, through an orchestration layer or meta-agent (Ren et al., 2025; Sapkota et al., 2025). These systems pursue goals rather than merely completing tasks. They decompose objectives into sub-goals, delegate subtasks among agents, integrate feedback from executions, and iteratively revise plans over extended timeframes. System-level intelligence arises from the coordination of components rather than from any individual element.

Agentic AI is characterized by two primary architectural features:

1. Structured memory. Agentic systems maintain both episodic memory, which records events from previous steps, and semantic memory, which stores long-term facts, vector representations, and tool schemas. These memory structures support continuity, reflection, error correction, and long-horizon reasoning, enabling the system to adapt as tasks evolve (Sapkota et al., 2025; Q. Zhang et al., 2025).

2. Orchestrated tool use and safeguards. Given that agentic systems can autonomously initiate tool calls, modify files, or perform operations with real-world consequences, deployments implement least-privilege access, approval checkpoints, tamper-resistant logs, and continuous monitoring to ensure safe execution (National Institute of Standards and Technology, 2024a).

Overall, the transition from GenAI to agents to Agentic AI can be understood along three axes emphasized in recent surveys: (1) *memory*: none $\rightarrow$ working memory $\rightarrow$ persistent episodic and semantic memory (Sapkota et al., 2025); (2) *planning*: single-step response $\rightarrow$ iterative task-level planning $\rightarrow$ continuous, goal-directed replanning (Ren et al., 2025; Wang et al., 2024); (3) *tool use*: optional retrieval $\rightarrow$ active API/function calling $\rightarrow$ orchestrated multi-agent toolchains (Piccialli et al., 2025). This progression is illustrated in Figure 2.3.

This taxonomy reflects prevailing enterprise workflows. Generative systems deliver single-step outputs under close supervision. Agents autonomously complete bounded tasks with moderate supervision. Agentic AI systems are designed for minimal supervision while maintaining policy compliance, auditability, and human override (Emenike, 2025; IBM, n.d.-a). As organizations integrate models with business systems and approval mechanisms, operational architectures are shifting from prompt-based interactions to outcome-driven, policy-constrained processes that span from user intent to task resolution (SAP, 2025a).

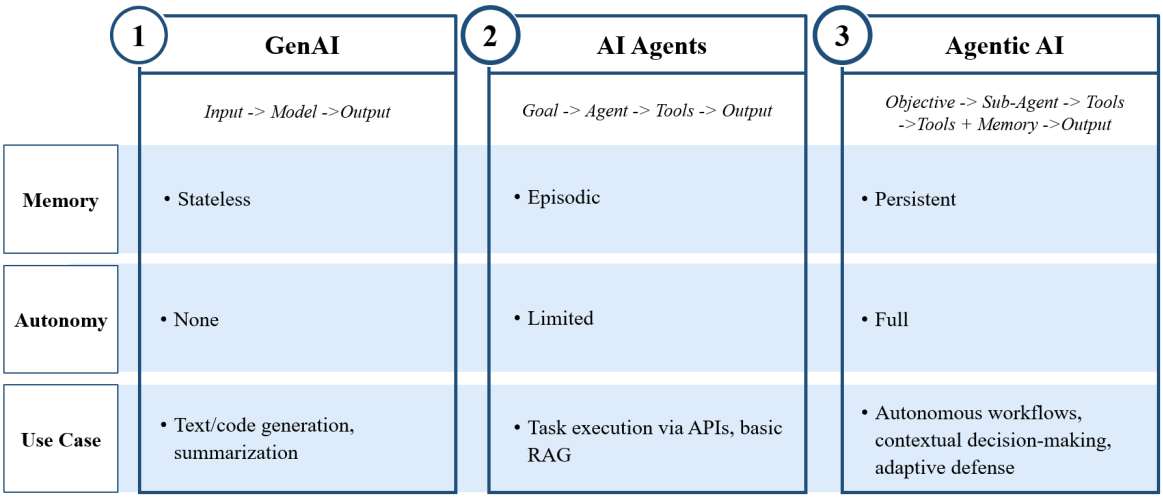


Figure 2.3: Architectural progression from GenAI to task-level agents and Agentic AI systems, adapted from Balbix (n.d.).

2.7. Agents as the Unifying Concept

Within this thesis, the term 'agent' refers specifically to the contemporary tool-using large language model (LLM) agent pattern, which executes bounded tasks through iterative planning and function or API calls. The GenAI → agent → Agentic AI taxonomy is therefore presented as an enterprise implementation progression of autonomy and orchestration, while remaining consistent with the broader classical agent framework.

The concept of an agent serves as a unifying framework within artificial intelligence. An agent is defined as any entity that perceives its environment through sensors and acts upon that environment through actuators (Russell & Norvig, 2021). In this context, the primary objective of AI is to design and analyze intelligent agents that map percepts to actions in ways that maximize performance relative to specified goals. The abstract agent function, which maps percept sequences to actions, is operationalized by an agent program running on a software or physical architecture equipped with appropriate sensing and actuation capabilities.

An intelligent agent is characterized as a rational agent, selecting actions that achieve the best expected outcome based on the information available. Rationality is determined by the agent's ability to maximize its performance measure under uncertainty, while considering computational constraints. A central aspect of rational agency is autonomy: instead of relying solely on fixed knowledge provided at design time, an intelligent agent must learn from experience to correct incomplete assumptions and adapt to dynamic environments (Russell & Norvig, 2021).

Modern AI Agents and Agentic AI in Practice

Large language models (LLMs) have revitalised the agent paradigm by enabling systems that integrate advanced reasoning capabilities with programmable interfaces. Contemporary AI agents operationalize classical agent concepts in software: perception is achieved through text or multimodal inputs, reasoning is conducted by LLMs, and actions are carried out via function calls, APIs, code execution, or retrieval pipelines. These systems maintain working memory and iteratively refine their plans based on observations, thereby aligning with model-based or goal-based agent architectures (Piccialli et al., 2025; Wang et al., 2024).

Agentic AI extends this paradigm to multi-agent, orchestrated systems in which specialized agents collaborate under a meta-agent or orchestration layer (Ren et al., 2025; Sapkota et al., 2025). This structure closely aligns with classical theory: agents maintain internal representations of the environment, decompose objectives into sub-goals, evaluate alternative actions, and enhance their behavior over time by updating memory structures, retrieval indexes, or tool usage. Thus, Agentic AI represents not a departure from classical agent theory, but rather its instantiation at unprecedented scale and complexity, employing LLMs as reasoning engines, APIs as actuators, and memory stores as state representations.

Recent applications demonstrate how these patterns manifest in practice. Process-planning agents translate high-level manufacturing specifications into detailed production steps (Holland & Chaudhari, 2024); decision-support agents in industrial contexts query enterprise data to inform carbon-management actions (Wu et al., 2024); and systems-management agents autonomously operate Linux servers using constrained command sequences (Cao et al., 2024). In all these domains, governance safeguards such as approvals, access control, and logging are essential to prevent cascading errors. Collectively,

modern agents and Agentic AI systems can be viewed as large-scale, LLM-enabled instantiations of classical rational-agent theory. They operationalize perception, reasoning, memory, and action within orchestrated digital environments, thereby extending the original agent taxonomy toward more capable, multi-agent, and goal-directed behavior.

2.8. Research Gap

The existing literature establishes a robust foundation for analyzing supply chain resilience and connecting it to operational decision-making through established conceptual frameworks and measurement perspectives. Resilience research has evolved from a narrow emphasis on preventing known risks to a broader capability-oriented perspective that links preparedness, absorption, recovery, and adaptation to key levers such as visibility, agility, flexibility, redundancy, and collaboration. Building on this foundation, an expanding body of research demonstrates that digital technologies, analytics, and artificial intelligence can enhance these levers, provided that data quality, integration, and governance are established. However, most existing evidence continues to reflect either traditional AI for prediction and optimization, or GenAI (GenAI), which is primarily utilized as an interpretive support layer for summarization and scenario ideation. As these systems move closer to operational decisions, their limits become binding.

As organizations begin to explore Agentic AI systems capable of executing multi-step workflows across enterprise tools, the supply chain resilience literature lacks actionable guidance on designing and governing such autonomy to enhance resilience rather than introduce additional complexity. Practitioner accounts suggest that Agentic AI can reduce response times; however, the necessary preconditions and governance mechanisms for consistently improving supply chain resilience are not well defined. The literature also remains unclear about the circumstances under which Agentic AI is justified for improving supply chain resilience, particularly given that many organizations continue to operate with fragmented data and partial system integration. Furthermore, the literature provides limited operational guidance on maintaining autonomous workflows safe under disruption conditions, which the Agent hasn't seen, where rare events can trigger cascading errors and time pressure reduces the margin available for human intervention.

A particularly underexplored and potentially high-value contribution involves translating safety-critical automation principles into design patterns for Agentic AI within supply chain decision-making. In aviation, automation is conceptualized not as a single capability but as a layered control system, in which distinct protections address various failure modes, and the human role is intentionally designed rather than assumed. Applying the principles that ensure safe automation in high-risk situations to Agentic AI can inform the development of best practices for deploying Agentic AI. However, the supply chain AI literature has not systematically adapted these aviation practices into actionable governance and training patterns for enterprise agents, nor demonstrated how these adaptations can reduce cascading errors while retaining the advantages that motivate Agentic AI autonomy. Additionally, the current literature does not offer a Supply Chain Operations Reference (SCOR)-aligned decision-point framework that specifies the appropriate level of autonomy for each supply chain decision according to its impact on resilience. Consequently, organizations lack systematic criteria for determining when GenAI should function as decision support and when Agentic AI should be granted execution authority to enhance supply chain resilience.

These gaps motivate the research questions:

Research Question (RQ).
How should organizations design and govern Agentic AI systems to improve supply chain resilience?

Sub-questions (SQs).

SQ1: *What data and system conditions justify choosing GenAI versus Agentic AI?*

SQ2: *What safeguards and human-in-the-loop controls are required to ensure safe and resilient autonomous operation?*

SQ3: *What training approaches improve an agent's robustness to unexpected supply chain events?*

SQ4: *Which mission-critical aviation automation principles can be translated into practical design patterns for Agentic AI to improve supply chain resilience?*

3

Research Method

Building on the literature review, this methodology is grounded in two main fundamentals: the Supply Chain Operations Reference (SCOR) model, which provides the base process language for the study, and a definition of supply chain resilience as the ability to maintain a firm's operational continuity despite uncertainties, risks, disruptions, and disturbances (Ivanov, 2021b). These two elements structure how the empirical analysis is designed and conducted.

3.1. Design Thinking behind the Research Methodology

This research follows a structured two-iteration process, with each iteration moving from divergence, where possibilities are explored, to convergence, where solutions are selected and specified in a disciplined way. The study therefore applies a deliberate diverge–converge rhythm: it first expands the space of options to surface a broad range of candidate interventions, and then narrows these options through structured evaluation to retain only those that are both novel and feasible. This sequencing is grounded in design practice and creativity research, showing that combining structured divergence with structured convergence improves both novelty and feasibility (Liu, 2017; Moreira, 2020).

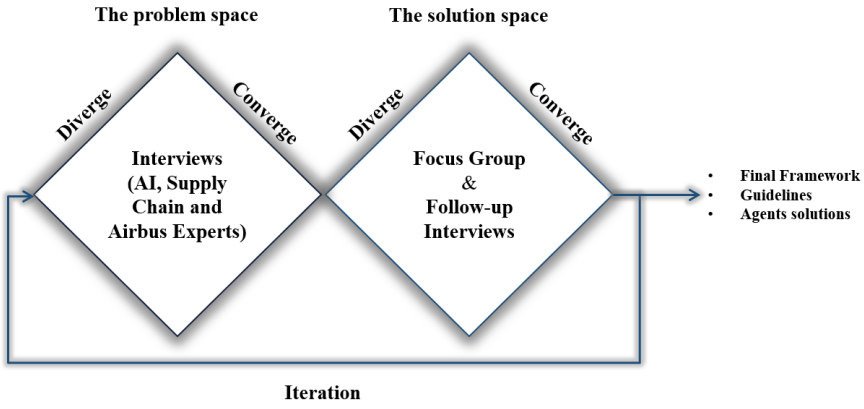


Figure 3.1: Double diamond adapted from Council (2025)

Table 3.1: Methods overview across research iterations

Iter.	Method	Participants	Purpose	Output	n_{part}
1	Semi-structured interviews	AI experts and supply chain experts	Understanding the actual potentials and functions of Agentic AI and the current problems in today's supply chains.	Initial framework	n_{part}=6
1	Focus group 1 and Follow-up interview	AI experts and supply chain experts	Co-design Agent concepts and delve deeper into a specific Agent use case.	Initial Agent portfolio and clear example of deployment in an industrial setting	n_{part} = 3
2	Semi-structured interviews	AI experts, supply chain experts and avionics	Test guidelines and delve deeper into specific concepts of AI. It also includes best practices from an avionics industry that uses automation in high-risk situations.	Validated guidelines, refined Agents and 8 pillars	n_{part} = 6
2	Focus group 2 and Follow-up interviews	AI experts and supply chain experts	Prioritize Agents and explore training approaches for out-of-ordinary events. Delve deeper into the training strategies used in the industry.	Final Agent selection and final strategy to train Agentic AI for unexpected events	n_{part} = 4

Concretely, the research: (i) explores the problem space through expert interviews and focus groups, (ii) synthesizes the findings into design options, (iii) converges on the best possible applications of Agentic AI and GenAI, and (iv) develops specific guidelines and best practices to identify where Agentic AI or GenAI applications are a good fit to improve supply chain resilience, derived from the interviews, and further informed by insights from Aviation experts.

First Iteration

In this phase, the interviewees are AI and supply chain experts, which makes it possible to develop a clearer understanding of the actual potentials and functions of these technologies and the current problems in today's supply chains. This is the divergent phase in which all the insights are gathered to create a solid foundation for the study. Working this way operationalizes the design-thinking, essential for finding the right problem before committing to any solution (Liu, 2017). Afterwards, the convergent phase starts, in which all the interviews are analyzed to gather all the relevant information and start developing specific guidelines.

The second part of the research reopens the lens in the *solution space*, but only within the boundaries established by the problem statements. In this phase, a focus group is used to analyze potential Agent concepts by combining the perspectives of supply chain experts and AI specialists. The discussion is then used to develop first drafts of Agents with clearly defined boundaries and rollback measures. As discussed in Section 2, each Agent, and the supply chain problem it addresses, is mapped to the SCOR framework. Each promising idea is captured in the same format to enforce comparability: the users and workflow it serves, the required data and systems, the expected effect on resilience metrics, the relevant risks and safeguards, and the smallest meaningful next step. These summaries shift evaluation away from pitch quality and toward decision-ready content (Sinfield et al., 2013; Wilson, 2013). To conclude, the output of this first iteration is the development of initial guidelines, first concept ideas of Agentic AI solutions and the initial pillars.

Second Iteration

The second iteration begins by validating the findings of the first iteration through targeted interviews, in order to assess the guidelines derived and compensate for specific aspects that did not emerge in previous interviews. Moreover, each agent derived is evaluated based on the impact on the dimensions highlighted in Section 2.1. However, it is crucial to highlight how this phase is still considered diverging, given the development of new agents based on the ones derived from the focus group. In this phase, these findings are then integrated with best practices in automation and AI in the aviation industry to identify how they can inform and enhance the deployment of Agentic AI for supply chain resilience. After analyzing and validating all possible ideas, the convergence phase begins, in which the agents are evaluated in the focus group, and the best use cases are selected. Moreover, the focus group explored which training techniques can help Agentic AI systems handle unexpected events. Two guardrails keep the method practical throughout. Each stage sets clear boundaries about what is in scope now, so that constraints focus creativity on feasible options rather than scattering attention on ideas that cannot be acted upon (Sinfield et al., 2013). In addition, ideas are made comparable before prioritization, by requiring the same fields on every summary, reducing bias toward the most charismatic speaker, and increasing the odds that the most testable options advance (Wilson, 2013).

In practical terms, the approach establishes a repeatable backbone for the study. Expert interviews diverge to understand and converge to define; possible applications diverge to invent and then converge to select the most promising ones. Following this cadence, the research turns qualitative inquiry into a sequence of evidence-seeking steps with a credible path to resilience impact.

End User

The end user of the design-thinking approach is the practitioner who must turn Agentic AI ideas into real, safe, and workable decisions in a supply chain. This includes supply chain leaders and process owners in operations, planning, logistics, and procurement, as well as consultants and transformation teams, who support prioritization and deployment across organizations. The diverge-converge rhythm is therefore used as a practical discipline rather than a creativity exercise. Divergence helps the research surface a broad set of resilience problems and possible AI applications from different expert perspectives. Convergence then forces structure and choice by making ideas comparable and selecting only those that are both feasible and defensible under real constraints.

3.2. Participants of Focus Group and Interviews

A diverse group of participants was selected to minimize bias. In the initial round of interviews, six experts participated. The first three were AI specialists: a professor with expertise in AI for mobility, a consultant experienced in AI and data-driven company strategy, and a data analyst specializing in machine learning techniques. The supply chain experts included a professor with a background in logistics and freight transport, a consultant specializing in supply chain strategy, and an availability leader from the fast-moving consumer goods (FMCG) industry, responsible for supply chain optimization and stock availability. The first focus group comprised a supply chain expert from a fresh food delivery company, an FMCG company employee, and the same data analyst, who also chose to participate in the focus group.

The second iteration included a professor specializing in manufacturing and supply chain systems, as well as a consultant focused on AI and machine learning applications in transportation and supply chain management. Additionally, an AI professor with expertise in explainability and a consultant experienced in AI applications within supply chains participated. The second focus group consisted exclusively of consultants with varying roles and seniority: a partner with over 14 years of experience in data science, machine learning, statistical data analysis, and AI; an AI consultant with 8 years of experience in AI and machine learning; and two supply chain experts, each with 8 years of experience. Furthermore, two interviews at Airbus were conducted with an AI and automation engineer possessing more than 30 years of experience and an Airbus pilot with over 20 years of experience.

3.3. Qualitative Analysis Interviews

Building on the backbone outlined in Section 3.1, the qualitative component operationalizes the divergent-convergent rhythm through a variety of interview rounds. For each round, it is crucial to follow specific steps: data gathering, data preparation, and qualitative analysis.

Qualitative Research Rationale and Inquiry Model

According to Lim et al. (2024), there are four principal reasons to undertake qualitative research: (i) to address complex social phenomena, (ii) to generate rich insights and human-centered understanding, (iii) to connect research to real-world issues, and (iv) to respond urgently to rapid social change. This thesis fulfills all of these criteria. The deployment here examined is a complex phenomenon with direct societal impact; therefore, it requires the in-depth, human-centered information that qualitative methods provide in order to establish where GenAI and Agentic AI are most suitable and likely to be effective. Moreover, much of the existing literature discusses the pros and cons of AI implementation without clarifying where, in a company's processes, it should be introduced. Finally, there is urgency to this work: many organizations implementing AI are not achieving their expected results, which reinforces the need for research that produces practical, actionable outputs.

Once it is established that this study falls within the domain of qualitative research, selecting an appropriate model to guide its execution becomes essential. The General Inquiry method, as introduced by Lim et al. (2024), is employed in this research. In contrast to grounded theory, which aims to develop new theoretical frameworks, General Inquiry is intended to explore a broad range of questions using diverse techniques, without imposing a predetermined theoretical perspective. As shown in Fig. 3.2, this

model structures the research process from the formulation of research questions to analysis, providing the methodological foundation for the data collection and thematic analysis procedures described below.

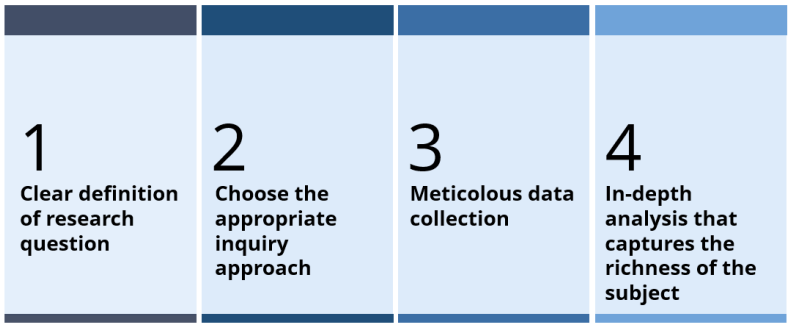


Figure 3.2: Process from RQs to analysis retrieved from W. M. Lim (2025)

Data Collection and Preparation

As highlighted in Fig. 3.2, step 3 regards the meticulous data collection. In this step is important to follow established protocols to ensure consistency in the various interview data preparations and to minimize potential biases. Data gathering precedes every type of analysis, and in this research it is carried out through semi-structured interviews with a diverse group of AI, supply chain, and aviation experts. The combination of multiple expert groups constitutes a form of data-source triangulation (Olsen, 2004): converging evidence a variety of stakeholders enhances the validity and credibility of the findings while keeping a coherent, qualitative strategy. In particular, this mixed-audience approach enables a robust analysis by incorporating diverse perspectives on capabilities, feasibility, human-AI collaboration, and governance, all of which are central to resilient operations. As in any inquiry that hinges on expert knowledge, it is crucial to identify stakeholders who are most relevant and possess the requisite domain expertise. Relevant stakeholders for this study are AI technical roles and supply chain roles. After identifying the roles, potential interviewees are contacted and interviews are scheduled.

Once collected, the audio recordings are processed using a standardized transcription protocol adapted from Pogna et al. (2025), ensuring that each interview is treated identically to avoid any possible bias. The process is designed to produce reliable, accurate, and analytically usable transcripts, as depicted in Fig. 3.3, which then serve as the direct input to the thematic analysis described in the following section.

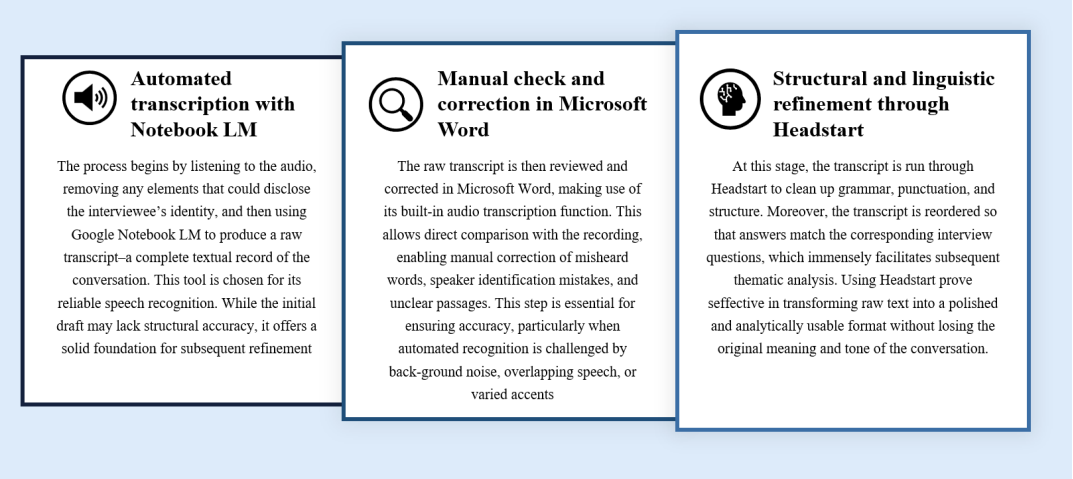


Figure 3.3: Data preparation adapted from (Pogna et al., 2025)

Analysis Procedures Across Iterations

Following the preparation of transcripts, the in-depth qualitative analysis begins. Given that interviews are conducted across both iterations, it is necessary to employ a consistent and rigorous analytical approach applicable throughout the study. Accordingly, the six-step thematic analysis framework proposed by Braun and Clarke (2006) is utilized. As Braun and Clarke (2006) described, is a structured way to analyze qualitative data. It proceeds through clear steps, as depicted in Fig.3.4. It begins with familiarizing yourself with the material, transcribing if necessary, reading and re-reading, and noting early ideas. The next step is coding, tagging meaningful features across the data set, and grouping related extracts. These codes are then collated into broader, potential themes. Themes are reviewed at two levels: first against the coded extracts, then against the entire data set, to ensure coherence and distinction. The process is recursive, with movement back and forth between data, codes, and themes, rather than strictly linear. Once themes are validated, they are refined, clearly defined, and named.

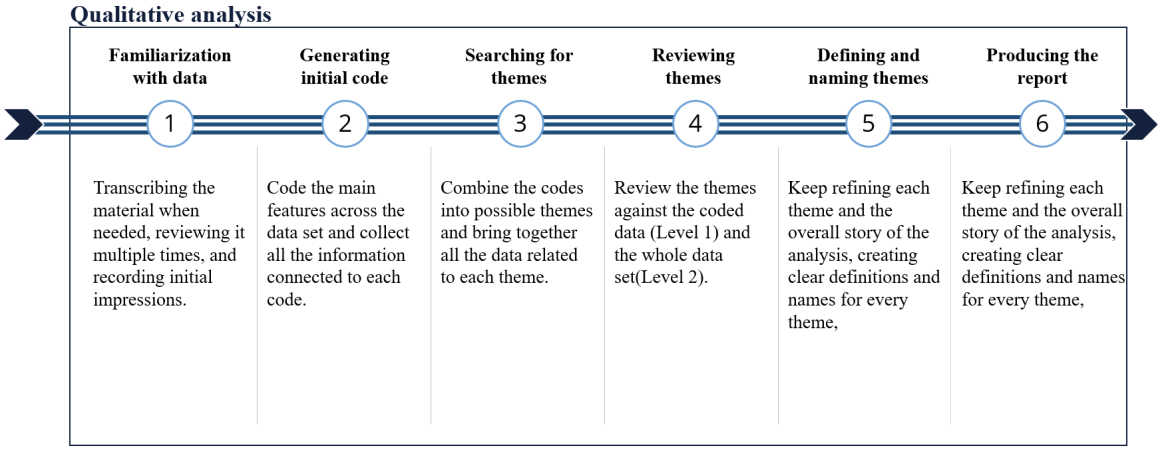


Figure 3.4: Six-step thematic analysis retrieved from Braun and Clarke (2006).

3.4. Focus Group Methodology

Methodological Background

Focus groups had their roots in the 1940s, when Robert K. Merton and Patricia L. Kendall introduced what they called the “focused interview”. Only later did this format come to be known as the “focus group”, marking its recognition as a distinct qualitative method. The core idea remained the same: a structured discussion led by a researcher with a small group of participants, who had all been exposed to a situation or topic that the researcher had already analyzed (Corrao, 2002; Merton & Kendall, 1946). In Merton and Kendall’s original formulation, the focused group interview was designed to elicit participants’ reactions to a previously defined situation, in order to understand how they interpreted it and which aspects they found most salient (Merton & Kendall, 1946).

Across the methodological literature, focus groups were described as especially useful in exploratory phases of research, when there was still limited systematic knowledge on a topic. In such contexts, they offered a flexible way to surface a broad range of experiences, interpretations, and problem framings that did not yet fit into fixed questionnaires or narrowly defined hypotheses (D. W. Stewart & Shamdasani, 1990; Wilkinson, 1998). A key feature was that the primary aim was not simply to collect individual answers, but to observe and analyze interaction between participants. As Bloor et al. (2001) emphasized, the value of focus groups was to stimulate discussion so that social norms, shared meanings, and disagreements became visible. In contrast to a standard group interview, where the interviewer mainly sought separate responses from each person, in a focus group, the moderator worked to encourage the processes of agreement, disagreement, negotiation, and joint sense-making, and these interactional processes were treated as the main data (Bloor et al., 2001; Parker & Tritter, 2006).

In definitional terms, a focus group was described as a research technique that collected data through group interactions on a topic set by the researcher. This highlighted three essential elements. First, the group was explicitly convened for data collection. Second, the data were derived from interactions within the group, rather than from isolated individual statements. Third, the discussion was actively constructed and guided by the researcher for clearly defined research purposes. This also clarified what was *not* a focus group: therapeutic or educational groups, decision-making meetings, or organizing committees if systematic data collection was not their main aim; methods that involved several people but no real interaction, such as nominal groups or Delphi techniques; and naturally occurring conversations without a moderator, because they lacked the research-driven structure (Morgan, 1996).

Focus groups were often combined with other qualitative methods, especially in-depth individual interviews. Empirical studies reviews showed that one of the most frequent designs combined focus groups and individual interviews within the same project (Morgan, 1996). The methods were seen as complementary: individual interviews allowed more detailed exploration of a person’s own experiences and reasoning, while focus groups revealed the range of opinions in a group and how these were shaped, challenged, or reinforced through interaction (Crabtree et al., 1993; Morgan, 1996). A common strategy was to use focus groups to identify a broad spectrum of experiences and perspectives, and then follow up with interviews to deepen the understanding of selected issues.

The composition and size of focus groups were also important design choices. Focus groups were organized to bring together different types of stakeholders, such as members of interest groups, NGOs, policy bodies, or practitioners from different organizational functions. In these cases, participants were invited in their specific roles and in relation to their particular stakes in the issue under discussion

(Swartling, 2007). Kontio et al. (2008) suggested that focus groups typically included between 3 and 12 participants, noting that smaller groups tended to show higher levels of involvement of each participant, whereas larger groups were associated with lower levels of individual participation. The moderator's role was to keep the discussion aligned with the research aims, use a clear sequence of questions, avoid dominance by a few participants, and maintain a permissive and respectful atmosphere that supported open and productive interaction (Parker & Tritter, 2006; Swartling, 2007). Conventional focus groups thus relied on what participants said and did in the group, as elicited and structured through this moderated interaction.

Finally, the literature highlighted three main ways in which focus groups had been used: (i) as a supplement to other methods in mixed-method designs, (ii) as a primary method in studies that explored how people experienced and interpreted issues, and (iii) as a tool in participatory action research to support learning and change (Wilkinson, 1998). This study adopted the first and second of these roles: focus groups were combined with semi-structured interviews in a mixed-method design, and served as a primary vehicle to explore how experts experienced and interpreted the use of Agentic AI for supply chain resilience.

Focus Group Analysis

Focus group transcripts are analyzed using the same qualitative procedure as the semi-structured interviews to ensure analytic consistency across data sources. Following transcription and familiarization, the data undergo open coding. These initial codes are then consolidated through thematic analysis, following the iterative steps outlined by Braun and Clarke (2006): codes are collated into candidate themes, themes are reviewed against coded extracts, and subsequently refined. The complete open-coding outputs and the resulting thematic structures are presented in Appendix B and Appendix D, respectively.

First Focus Group: Exploration of Resilience Problems and Agents

The first focus group was used in an exploratory way to identify concrete resilience problems along the SCOR chain, to co-design candidate Agents to address these problems, and to define initial guardrails for their use. The group included three experts, which was in line with Kontio et al. (2008), who noted that small groups of around three participants allowed a high level of involvement of each person. The stakeholders were selected for their complementary expertise: two participants with supply chain experience and one with AI experience. This composition reflected the recommendation by Gerger Swartling (2007) that focus groups should bring together different types of stakeholders relevant to the issue.

The session was moderated by the researcher and structured around concrete disruption scenarios and pain points mapped to the SCOR processes. Rather than simply asking each participant for individual opinions, the moderator encouraged interaction - agreement, disagreement, and negotiation - so that problems and potential Agent concepts emerged from collective sense-making rather than from isolated statements (Bloor et al., 2001; Parker & Tritter, 2006). Because the focus group was conducted online, interactive boards were used to support collaboration. From the third step onwards, ideas were rotated so that each participant could add to the others' inputs, suggest mitigation plans, and comment on feasibility. This design directly reflected the aim described by Bloor et al. (2001), namely to stimulate discussion and to make visible how shared and divergent interpretations developed in the group.

Agentic Claims Processing Case Study

After the initial focus group, a participant implemented an agent concept that had emerged during the session. Due to the practical significance of this implementation, a dedicated case study was conducted to examine it in detail. A single-case study design was selected, which is appropriate for investigating contemporary phenomena within real-life contexts (Yin, 2003). The case was chosen through purposeful sampling (Patton, 2014) as an information-rich critical case, allowing for logical generalization about the transition from deterministic batch processing to agent deployment. Data collection involved an in-depth, semi-structured interview (Patton, 2014), which enabled the practitioner responsible for workflow design and implementation to explain their perspective and decision-making rationale. The interview data were corroborated with technical documentation, such as workflow diagrams and process specifications. The transcript was analyzed also in this case using the approach described by Braun and Clarke (2006).

Second Focus Group: Validation of Agents and Guidelines

Building on the results of the first focus group and on subsequent expert interviews, a second focus group was organized with a mixed group of AI and supply chain experts. The aim of this second session was not to generate new ideas from scratch, but to validate and refine the Agent concepts that had emerged, and to delve deeper into possible strategies to improve the training of AI systems. In this way, the second focus group moved the research from a purely exploratory phase toward a more evaluative and convergent phase, while still leveraging interaction as the primary data source (Morgan, 1996; Wilkinson, 1998).

More specifically, the second focus group had two main objectives. First, the groups was given the entire list of Agents and asked to identify the best ones in terms of AI feasibility and supply chain impact, and to comment on necessary improvements, and to propose possible additional or derived Agents. Second, the group was asked to reflect on how to correctly prepare AI models to deal with unexpected events.

As in the first focus group, the session was moderated by the researcher and was structured into four parts. In Part 1, participants briefly introduced themselves and their backgrounds, and received a short reminder of the research goal and the working definitions of GenAI and Agentic AI. In Part 2, they worked in pairs with a list of Agents, derived from the first focus group and the interviews. For each Agent, they discussed current feasibility, whether it should be split into more focused Agents or merged with overlapping ones, and then they selected the ten Agents they considered most important, explaining their choices and suggesting improvements or new complementary Agents.

In Part 3, the pairs came together in plenary to compare their selections and jointly agree on the five most relevant Agents overall, discussing trade-offs and prioritization criteria. Finally, in Part 4, an open group discussion was used to delve deeper into possible ways to train GenAI and Agentic AI for unexpected events outside of the ordinary. The second focus group thus provided as a structured, interaction-centered setting where participants collectively scrutinised both the Agent concepts and the technology improvements needed to handle unexpected events, before these were forward used as input to the subsequent research stages.

3.5. From Aviation to Supply Chains: A Cross-Industry Analogy for AI Governance

This study applies cross-industry analogical reasoning to transfer validated governance practices from a source industry to the context of supply-chain decision making. Rather than asserting equivalence between industries, the objective is to identify structural similarities such as high-stakes decisions under uncertainty, the consequences of error, and the need for oversight, and to translate the underlying principles that render practices effective in the source domain. Cognitive science demonstrates that robust analogies rely on mapping relational and causal structures rather than surface features, thereby reducing the risk of importing irrelevant attributes (Gentner, 1983). Research on design-by-analogy similarly highlights the importance of abstracting transferable mechanisms instead of replicating literal solutions (Christensen & Schunn, 2007; Goel, 2002). In contrast to intra-industry benchmarking, this study adopts a Cross-Industry Innovation approach, in which practices from different industries enable ”creative imitation” through abstraction and adaptation (Enkel & Gassmann, 2010). The selection of analogies is based on Structure-Mapping Theory (SMT), which considers structural correspondence as the primary criterion for validity (Gentner, 1983).

The Aviation industry is selected as the comparative domain because the partially automated process of flying an aircraft — in which authority is continuously allocated between pilot and system under codified governance rules captures the kind of advanced governance needed for AI-enabled supply-chain decision support. In both sectors, effective automation depends on clearly defined decision roles and escalation logic, which allow authority to shift appropriately as uncertainty increases or performance declines. Both domains utilize mechanisms for issue detection and prioritization that filter and rank deviations, thereby directing limited human attention during periods of information overload. When components fail or data quality deteriorates, explicit fallback and recovery modes are required to reduce complexity and maintain safe operation during diagnosis. These arrangements are effective only if system-state transparency is maintained, ensuring that modes and system status are communicated clearly to minimize confusion and prevent inappropriate reliance on automation. Additionally, both domains benefit from standard operating procedures that coordinate responses through shared protocols, sustained by organizational learning rather than ad hoc improvisation. Continuous improvement is supported by post-event review and traceability, as decisions and outcomes are logged and analyzed to refine governance based on evidence rather than relying solely on real-time adaptation.

The analogy remains valid despite differences in physical domain, timescale, and regulatory context because it focuses on transferable governance patterns rather than shared physical substrates or identical environments. Aviation operates within a bounded flight envelope with rapid consequences, whereas supply chains manage distributed flows in which deviations appear as planning mismatches. Aviation is subject to stringent safety regulations and mandatory reporting, while supply chains are primarily driven by commercial incentives and generally lack comparable certification and incident-reporting systems. These differences establish boundary conditions for translation but do not undermine the core correspondence: both domains must determine acceptable levels of autonomy, specify when autonomy should be reduced, and ensure that humans can supervise, understand, and override automation when reliability decreases.

4

First Iteration Interviews - AI and Supply Chain Experts

This chapter reports the findings from the first interview iteration of the study, conducted within the first cycle of the double diamond methodology. The purpose of this iteration is to establish a robust empirical baseline by collecting expert perspectives from two domains: AI (to clarify what Agentic AI and GenAI can realistically do, and under which safeguards) and supply chain resilience (to understand the problems they face on a daily-basis). In this way, the interviews connect capability and governance on the AI side with decision points and disruption realities on the supply chain side, creating the foundation for later convergence into design principles and the framework.

Methodologically, this chapter applies a structured qualitative analysis pipeline to improve transparency and reduce interpretive bias. As illustrated in Figure 3.2, each interview is processed through a consistent sequence of data preparation, open coding and thematic analysis. For each interview, the results are documented in a thematic analysis table that consolidates the main themes and the codes supporting them. For transparency and replicability, the summary of interview transcripts are provided in the Appendix, together with the complete open-coding tables used as the basis for themes development.

The chapter is organized in two main parts. The first part analyzes three interviews with AI experts (consultancy, academia, and data science), focusing on definitions and boundaries of Agentic AI, typical failure modes, and the governance and oversight mechanisms required for safe deployment. The second part analyzes three supply chain expert interviews (industry, consultancy, and academia), focusing on disruption types and propagation, resilience policies and mitigation strategies, and the organizational and data conditions that determine whether technology can be adopted in practice. The chapter then synthesizes each part into macro-categories of themes, from which the guidelines and the framework will be developed.

4.1. First Iteration Interviews - AI Experts

This section presents a thematic analysis of three expert interviews. Semi-structured interviews of 45 minutes each were conducted, recorded, transcribed, and analyzed. Initial open coding identified codes, consolidated into themes within each interview.

The following sections present: (i) individual interview analyzes with thematic tables, (ii) macro-category synthesis.

Table 4.1: Themes and Definitions – First Interview (AI Expert)

Theme Name	Definition
Theme 1: AI Basics and Agentic AI	This theme elucidates the conceptual positioning of Agentic AI within the broader artificial intelligence landscape. Artificial intelligence serves as the overarching category, encompassing machine learning, while GenAI and Agentic AI are identified as more specific capability families. The primary distinction is agency: Agentic AI is characterized by a closed-loop operational pattern in which the system sets goals or plans, executes actions - often utilizing tools - and subsequently evaluates and adjusts its outputs, rather than generating a single, one-time response.
Theme 2: Rules and Safety Controls	This theme addresses the necessary conditions for the responsible deployment of Agentic AI systems within organizations. It highlights that autonomy should be constrained by governance mechanisms, including policies, defined autonomy levels, access controls, and accountability structures, and must be reinforced by human oversight and human in the loop (HITL) as well as monitoring and control systems. Importantly, safety is not considered an additional feature; rather, it must be integrated from the outset through safeguards such as data protection, exception prevention, and mechanisms that facilitate post-incident learning, including root-cause analysis and traceability.
Theme 3: Where AI Goes Wrong	This theme examines the reasons for AI system failures and the mechanisms by which errors propagate in practical applications. It identifies two primary drivers of failure: hallucination, defined as credible but incorrect outputs, and poor inputs, often described as garbage-in/garbage-out. Notably, the analysis emphasizes that Agentic AI architectures do not eliminate errors. Even when an agent validates intermediate steps, the overall workflow may still converge to an incorrect result, sometimes including fabricated figures, particularly when inputs or tool outputs are unreliable.

Continued on next page.

Table 4.1 continued.

Theme Name	Definition
Theme 4: Costs and Benefits	This theme examines the economic rationale for adopting Agentic AI and addresses why it may not always represent the optimal choice. Value is defined by increased speed, enhanced quality, or greater scale of work, alongside organizational benefits such as improved employee utilization and job satisfaction through the reassignment of personnel from low-value tasks. However, Agentic AI generally incurs higher unit costs compared to GenAI due to the need for additional prompts, orchestration steps, tool integrations, monitoring, and, in some cases, multiple models, as well as greater initial implementation effort.
Theme 5: Readiness and Pilot Strategy	This theme outlines the organizational prerequisites necessary for successful adoption and the selection of scalable initial deployments. Two primary readiness conditions are identified: data maturity, which involves understanding data sources, definitions, and quality while avoiding unreliable inputs, and employee AI maturity, which refers to the skills and behaviors required to use AI effectively beyond basic tasks. The discussion also addresses pilot strategy, emphasizing that effective pilots should focus on end-to-end processes, particularly repetitive workflows where coordinated steps yield significant improvements. A key barrier highlighted is risk aversion, which impedes experimentation and restricts proof-of-concept initiatives.
Theme 6: Best Supply Chain Resiliency Uses	This theme examines the distinct value of Agentic AI in enhancing supply chain resiliency, particularly for tasks necessitating extensive cross-system coordination. Two principal application areas are emphasized: data reconciliation, which includes detecting missing or inconsistent records, harmonizing data formats, validating accuracy, updating authoritative systems, and engaging data owners for resolution; and proactive event or risk monitoring, which involves tracking weak signals from multiple sources, coordinating investigative actions, and managing follow-up activities. The central argument asserts that effective supply chain performance depends on orchestrating fragmented data and workflows, a process that aligns with the core capabilities of Agentic AI mechanisms.

Table 4.2: Themes and Definitions – Second Interview (AI Expert)

Theme Name	Definition
Theme 1: AI and Agentic AI Fundamentals	This section defines AI, GenAI, and Agentic AI, and describes their respective structures. AI represents the overarching field, while GenAI is a subfield focused on learning patterns in data to generate outputs, often exemplified by large language models (LLMs). Agentic AI extends GenAI by integrating multiple sequential steps and coordinating models or agents capable of utilizing external tools. The traffic-light example demonstrates an Agentic AI workflow in practice, which involves perception, optimization, and issuing actions to a controller.
Theme 2: Oversight and Autonomy	Oversight should be determined by the level of task risk and contextual factors, rather than solely by whether the system is classified as GenAI or Agentic AI. In situations with severe consequences, such as safety-critical applications like traffic control, strong human oversight is essential. For lower-risk tasks, including those that are simple or have demonstrated performance exceeding that of humans, greater autonomy may be appropriate. The degree of oversight should be established by comparing machine performance to that of competent human operators and adjusted according to the potential consequences. Notably, risk increases rapidly as the operational scope expands, for example, from managing a single intersection to overseeing an entire traffic network.
Theme 3: Failure Modes and Threats	This theme outlines primary failure modes in AI systems and explains how Agentic AI configurations can introduce additional risks. In generative AI, hallucination represents a significant concern. Broader risks encompass data integrity issues, such as polluted training data or memory poisoning, as well as bias originating from training data and input manipulation through prompt injection. Agentic AI systems introduce coordination risks, including the potential for agents to contradict one another, the possibility that a single compromised component may disrupt the workflow, and the likelihood of failures occurring during handoffs. Failure detection varies in difficulty; tool failures are generally more apparent, whereas plan drift during multi-step execution is more challenging to identify. Furthermore, increased model complexity reduces transparency, thereby making unexpected outcomes more difficult to anticipate.

Continued on next page.

Table 4.2 continued.

Theme Name	Definition
Theme 4: Safety Controls and Recovery	This theme addresses controls designed to enhance system safety and reliability. Prior to deployment, systems require systematic evaluation and iterative testing. Following deployment, continuous oversight is necessary, particularly in safety-critical applications. When explicit rules are difficult to define, mitigation strategies may involve a separate evaluator or critic agent and cross-validation, such as cross-agent critique or verification against a database. During operation, exceptions should prompt reduced autonomy, escalation to human operators, or the implementation of compensatory strategies. To support recovery and resilience, systems should maintain robust state logging for rollback, secure and backed-up memories and data, and well-defined playbooks for restoring operations.
Theme 5: Deployment Readiness and Governance	This theme addresses the organizational requirements and decision criteria relevant to adoption. Key needs include establishing a technical and safety function that translates standards into engineering checks, assigning clear legal responsibility for potential harm, and implementing robust governance over data and access controls, including authentication and permissioning. Evaluation of success should involve key performance indicators (KPIs) that assess both goal achievement and risk minimization, such as robustness, fairness, and transparency. Robustness checks should include adversarial testing and monitoring for distribution shifts. Additionally, Agentic AI systems are generally more costly due to higher computational demands, with expenses influenced by training methods, model size, and data volume. Consequently, pilot implementations should begin with low-risk, well-defined tasks that have a reference process, validate both performance and safeguards, and scale autonomy incrementally.

Table 4.3: Themes and Definitions – Third Interview (AI Expert)

Theme Name	Definition
Theme 1: AI, GenAI, and Agentic AI: Definitions and Scope	This section delineates three concepts according to their functional distinctions. Artificial intelligence (AI) encompasses a broad range of machine learning methods used for prediction and classification, including simple models. GenAI represents a narrower subset, typically involving large language models (LLMs) that generate text. Agentic AI extends these capabilities by integrating an LLM with external tools such as application programming interfaces (APIs), databases, files, and sensors to execute multi-step workflows with a degree of autonomy.

Continued on next page.

Table 4.3 continued.

Theme Name	Definition
Theme 2: Model Limitations and Operational Threats	This section outlines the primary limitations of large language models (LLMs) and the operational risks associated with deploying agents. LLM limitations include hallucination, inadequate performance in exact mathematical computations, restricted memory during extended sessions, and knowledge cut-offs. Operational risks encompass prompt injection, which involves overriding instructions, and potential data exposure during cloud-based operations. In Agentic AI workflows, these risks are amplified, as a single error can propagate and result in multiple subsequent failures, a phenomenon referred to as 'extreme agency.'
Theme 3: Oversight and Safeguards	Safe deployment requires aligning the level of autonomy with the associated risk and subjectivity of tasks. Tasks with higher risk or greater subjectivity necessitate reduced autonomy, enhanced controls, and increased human oversight. Greater control typically leads to improved precision. Key safeguards include understanding model limitations by chunking inputs and verifying outputs, incorporating deterministic steps for exact calculations, maintaining human involvement for judgment, and enforcing robust security measures such as protected credentials, approved models or providers, and controlled access. Organizational prerequisites involve acknowledging the possibility of failure, prioritizing preventive measures, maintaining strong cybersecurity, and implementing formal governance of access.
Theme 4: Exceptions and Recovery	This section outlines methods for agents to detect internal failures within the AI system itself and prevent damage. Two primary techniques are agency over agency, in which a secondary agent reviews decisions to identify anomalies or drift, and deterministic guards, which involve range, schema, or invariant checks. If these checks fail, the system should initiate hard stops and transfer control to human operators, enforce explicit thresholds to halt or revert changes, and implement rollback mechanisms to prevent error propagation. The objective is to ensure stable system behavior through graceful degradation, escalation, and safe compensation.

Continued on next page.

Table 4.3 continued.

Theme Name	Definition
Theme 5: Economics and Task Selection	This section outlines the criteria for suitability and the underlying business logic. Generative AI is most appropriate for tasks that require significant judgment or involve high stakes, where human oversight is essential for final decision-making. In contrast, Agentic AI is better suited for repetitive, low-creativity tasks that process large volumes of unstructured data and where automated analysis should directly trigger actions. Key performance indicators (KPIs) such as cost per use (measured in tokens) and latency should be used to assess success, with careful consideration of the trade-offs between quality, speed, and required accuracy. Operational costs increase with higher token usage and are further amplified by orchestration and infrastructure demands, frequent API calls, retries, and the compounding effect of "agency over agency," which escalates the number of calls and associated expenses. When selecting pilot projects, priority should be given to domains characterized by text-heavy content, repetitive processes, and a well-defined business case, particularly in scenarios requiring continuous monitoring.
Theme 6: Supply chain resilience applications	This section examines the benefits of Agentic AI systems in supply chains. Agentic AI systems are particularly effective in contexts that require reading, summarizing, and acting on documents, as well as in environments characterized by frequent and repetitive decision-making. These systems can transform review processes from periodic to daily by converting unstructured information into timely, actionable insights. To enhance resilience, agents can continuously monitor external signals such as news and maritime data, track indicators near critical chokepoints including ship density and speed, and recommend route or plan adjustments more rapidly than traditional periodic reviews. However, implementing these capabilities at scale necessitates significant engineering resources and incurs additional costs.

4.2. First Iteration Interviews - Supply Chain Experts

This section provides a thematic analysis of three expert interviews. The interviewees were selected from diverse supply chain backgrounds, including consultancy, academia, and fast-moving consumer goods (FMCG). Each semi-structured interview lasted 45 minutes and was subsequently recorded, transcribed, and analyzed. Initial open coding was used to identify and consolidate codes into themes within each interview. The following section details the thematic analysis.

Table 4.4: Themes and Definitions - First Interview (Supply Chain Expert)

Theme Name	Definition
Theme 1: Role Scope and Daily Workflow	This section outlines the responsibilities of the Availability Leader and the structure of the workday. The position involves both proactive monitoring, particularly during major promotions, and reactive interventions to address stock issues as they arise in the market. The daily workflow is highly structured: risks are assessed each morning, minor issues are disregarded, and significant risks are tracked using a central European risk management tool. Escalation occurs only when the potential impact is substantial, such as approximately ten pallets. Following the 11:20 order cut-off, the focus shifts to the release phase, during which limited stock is distributed among orders. Additionally, the role encompasses the management of non-performing and aged inventory in collaboration with the sales team.
Theme 2: Common Shocks and Failure Points	This theme categorizes the primary sources of operational instability. Frequently occurring issues include truck delays, late internal shuttles, and distribution center congestion or unloading constraints. Less common but high-impact disruptions involve infrastructure failures such as canal blockages, theft or returns resulting from security breaches, and unreliable rail service necessitating urgent transitions to truck transport. Additionally, this theme addresses demand shocks, such as unexpected changes in customer promotion mechanics that lead to demand surges exceeding the weekly plan.

Continued on next page.

Table 4.4 continued.

Theme Name	Definition
Theme 3: Crisis Response and Hard Constraints	This theme examines scenarios in which disruptions exceed the capacity of standard corrective measures. In such cases, the Availability Leader and planners must implement rapid mitigation strategies, such as negotiating delivery schedules with customers, reallocating production volumes across plants, and sourcing inventory from alternative markets or borrowing safety stock when permitted by labeling regulations. The analysis also highlights fixed constraints, including bottlenecks such as cardboard lead times and supply dependencies, which can extend recovery periods to several weeks. The typical 3–4 week lead time precludes stock recovery for short-term promotions following a demand spike. Decision-making in these situations requires balancing immediate promotional needs against the risk of future stockouts.
Theme 4: Resilience Levers and Inventory Governance	This theme addresses the inherent resilience of the system and the key performance indicators (KPIs) and rules implemented for its management. Structural resilience is achieved through a European multi-plant network, which enables rapid production shifts and minimizes the duration of disruptions, as demonstrated by the Ukraine case where impacts were limited to several weeks rather than becoming prolonged. Operational resilience is supported by promotion visibility and the capacity to reallocate volume across markets. The primary metric is IWL (the share of assortment not replenished to safety stock); when this metric exceeds the 5% threshold, service issues in the following week are considered highly probable. The theme further outlines explicit threshold rules for managing excess stock: scrapping for low-value items, brokering for medium-value items, and sales alignment for high-value items.
Theme 5: Organizational Friction and External Partnerships	This theme addresses the primary barriers related to people and systems, highlighting the contrast between internal and external collaboration. Internally, decision-making is often compromised by misaligned key performance indicators, as teams prioritize their own metrics, such as planners providing conservative estimates. This behavior generates friction and diminishes overall profitability. In contrast, external collaboration enhances resilience when customers participate in forecasting and provide order visibility. The interview further identifies incentive mechanisms, including rewards for full-truckload ordering to improve logistics efficiency and agreements in which customers place promotional orders earlier in return for guaranteed delivery. Additionally, the analysis notes constraints such as sustainability strategies that restrict rapid material changes and recurring local challenges, such as strikes, which permit only partial pre-planning.

Continued on next page.

Table 4.4 continued.

Theme Name	Definition
Theme 6: Internal AI Use and Future Agent Use	This section outlines current applications of artificial intelligence (AI) and potential future developments in automation. Presently, AI is primarily utilized for internal, non-operational purposes. In-house tools are implemented to maintain confidentiality and support learning and communication, such as decoding organizational terminology and training for managerial interactions. In operational contexts, the initial phase of automation involves rule-based systems, including the construction of decision trees to automate routine risk assessments and the standardization of excess-stock analysis using monetary thresholds. In the longer term, the envisioned approach involves AI copilots generating scenario proposals and options, while automated agents execute standard reallocations and rescheduling within defined parameters, such as financial thresholds and approval workflows. This approach is particularly suitable for planning functions, as these decisions are typically data-intensive and repetitive.

Table 4.5: Themes and Definitions - Second Interview (Supply Chain Expert)

Theme Name	Definition
Theme 1: Shock Drivers and Early Signals	This theme examines the primary forces that destabilized supply chains and the initial ways in which companies identified these disruptions. The COVID-19 pandemic resulted in halted flows, suspended deliveries, and significantly increased logistics costs, prompting strategic considerations such as nearshoring and dual sourcing. The introduction of new sustainability regulations generated a distinct shock, requiring firms to reassess suppliers associated with deforestation, state ownership, or compliance risks. Additionally, natural disasters frequently affect deep-tier suppliers (Tier-3/4 and beyond), with the resulting disruptions propagating downstream in a manner similar to the bullwhip effect.
Theme 2: Resilience Strategy and Governance	This theme addresses the foundational principles and governance approaches required to enhance the resilience of the upstream supply base. The interviewee conceptualizes resilience as a progression from Visibility to Transparency to Traceability (VTT), with traceability representing the stage at which organizations can proactively address upstream value-chain challenges. The initiatives discussed are primarily conceptual, focusing on governance and sourcing strategy rather than immediate operational responses. A significant barrier identified is the prevailing perception of procurement as primarily a cost-minimization function, which complicates efforts to communicate the strategic importance of resilience to senior management.

Continued on next page.

Table 4.5 continued.

Theme Name	Definition
Theme 3: Risk Detection Tools and Upstream Focus	This section examines technologies designed to enhance upstream visibility and risk identification. A key finding is the absence of a unified internal dashboard that proactively flags issues. The interviewee referenced an advanced AI-based risk tool that maps both tier-1 and tier-N risks, generating a prioritized list for management, primarily within indirect procurement. Most resilience tools emphasize upstream processes, such as visibility and traceability, rather than integrating both risk identification and operational response. Given that most disruptions originate upstream, prioritizing the SCOR ”Source” process for Agentic AI. For instance, upstream AI applications may combine geospatial data and image recognition to monitor deforestation risks.
Theme 4: Data Quality and the Data-Sharing Barrier	This theme highlights the challenges in ensuring the reliability of upstream risk tools, primarily due to incomplete, inconsistent, or poorly integrated supply chain data. As a result, many tools rely on open-source data and probabilistic estimates, such as inferring deep-tier relationships without comprehensive traceability. According to the interviewee, structured data sharing among partners represents the most effective means of enhancing resilience, as it increases the predictability of deep-tier exposure. Nevertheless, numerous companies are reluctant to share supplier data, perceiving it as a source of competitive advantage.

Table 4.6: Themes and Definitions - Third Interview (Supply Chain Expert)

Theme Name	Definition
Theme 1: Operational Focus and Risk Landscape	This theme establishes the operational context and identifies the primary sources of risk. The interviewee primarily operates within the SCOR Make and Deliver domains. The highlighted risks include localized network failures such as flooding that disrupts rail or road links, deteriorating infrastructure requiring significant investment, and increased traffic driven by internet ordering, which places additional pressure on logistics networks.
Theme 2: Reactive Detection and Low-Inventory Escalation	This theme describes a mostly reactive detection system. Issues are noticed when stock is physically missing during picking or when customers complain. The disruption then spreads quickly because low inventory levels mean any missing part can block production, turning a small shortage into a major production issue.

Continued on next page.

Table 4.6 continued.

Theme Name	Definition
Theme 3: Resilience Strategy and Collaboration	This theme presents practical resilience measures and emphasizes the importance of partnerships. The interviewee recommends maintaining higher safety stock as a buffer. Additionally, a traffic-light strategy (green, yellow, red) and risk classification of suppliers and customers are proposed, with multi-sourcing suggested for high-risk suppliers. Resilience is described as a process involving three steps: early identification of problems through strategic planning, internal alignment on solutions, and clear communication with external partners. Effective collaboration relies on open communication, regular meetings, and acknowledgment that the company itself may also contribute to challenges.
Theme 4: IT Monitoring, AI Use, and Safe Automation	This theme reflects the interviewee's perspective on technology. The interviewee expresses skepticism toward generic AI tools, citing insufficient data, lack of contextual understanding, and frequent disconnection among IT systems, which leads to data quality issues. They emphasize that organizational resilience requires a robust IT infrastructure with continuous monitoring and alert mechanisms. According to the interviewee, Agentic AI is primarily valuable for pattern recognition when supervised by humans. Automation should initially target simple, repetitive decisions, and a critical safeguard is to avoid automating mission-critical tasks without human oversight.

4.3. Concluding Remarks

This chapter synthesizes the insights from three AI experts (consultancy, academia, and data science) and three supply chain experts (industry, consultancy, and academia), and demonstrates that the central issue is not a binary choice between Agentic AI and non-Agentic AI. Instead, the focus should be on determining the appropriate level of automation for each decision point, given real-world constraints such as tight deadlines, incomplete data, conflicting incentives, and exposure to disruption.

The interviews with AI experts identified distinct capability boundaries. GenAI is most effective in generating human-verifiable advisory outputs, such as drafts, analyzes, and recommendations, where responsibility and decision rights remain with human operators. Agentic AI architectures are advantageous when tasks require the orchestration of multiple tools, data sources, and process steps through iterative plan-act-verify cycles, with the potential to execute actions under explicit safeguards. However, increased autonomy introduces more risks. In addition to model-level concerns such as hallucination, bias, and data quality, Agentic AI systems present system-level vulnerabilities, including cascading errors, privilege escalation through tool use, prompt injection, data leakage, and unsafe goal pursuit if objectives are mis-specified. All experts agreed on a core principle: autonomy must be explicitly calibrated to the associated risk level. Achieving this requires an integrated operating architecture that combines policy-layer governance (autonomy levels, approval points), technical safeguards (deterministic checks, constrained access), operational controls (monitoring, human-in-the-loop), and exception management (hard stops, escalation, rollback). The interviews also clarified economic boundaries.

GenAI typically delivers return on investment through productivity gains in judgment-based tasks, while Agentic AI approaches justify their higher operating and assurance costs when they can safely automate high-volume workflows or high-impact, time-critical response routines with robust controls. Initial deployments should focus on well-defined processes with clear business cases, particularly in supply chain applications such as continuous data reconciliation and external disruption monitoring.

The supply chain expert interviews indicated that disruption is a structural characteristic rather than an exception. Both frequent operational disturbances, such as transport delays and distribution center congestion, and rare high-impact shocks, including geopolitical events and material shortages, propagate rapidly through lean, tightly coupled systems when detection is reactive and buffers are minimal. Resilience is achieved through three interconnected layers that implement core capabilities: visibility, flexibility, redundancy, and collaboration. These layers include structural design (multi-plant networks, safety stocks), operational routines (daily cycles, cut-offs, playbooks), and governance mechanisms (risk classification, collaboration incentives). However, persistent frictions limit effectiveness. Misaligned key performance indicators encourage conservative behavior, procurement remains primarily cost-focused, IT environments are fragmented, and data sharing is hindered by confidentiality and competitive concerns. Current AI applications primarily serve as decision-support tools, informing prioritization and diagnosis rather than autonomously committing inventory, capacity, or routing changes. Progression toward Agentic AI support should be incremental, focusing on automating repetitive, rule-based tasks within clearly defined safeguards.

Analysis of both sets of interviews identifies six fundamental task families as the most suitable initial applications for Agentic AI in enhancing supply chain resilience. The first is *data cleaning and alignment*, which involves detecting missing or inconsistent records across systems, harmonizing data formats, validating accuracy against authoritative sources, and engaging data owners for resolution. The second is *proactive external monitoring*, which entails tracking weak signals from heterogeneous sources such as news feeds, maritime data, and geospatial imagery, particularly near critical chokepoints, and triggering predefined investigative or escalation actions. The third is *daily risk briefing*, which automates the daily morning cycle of risk assessment by scanning inventory positions, transport statuses, and order books to pre-prioritize issues before human review. The fourth is *excess and non-performing stock disposal*, which applies monetary and volume thresholds to classify aged inventory and generate disposition recommendations, such as scrapping, brokering, or sales alignment, within approved rules. The fifth is *document screening*, which consists of reading, summarizing, and acting on compliance documents, supplier assessments, and regulatory updates, thereby converting periodic review processes into near-continuous ones. The sixth is *standard reallocation and rescheduling*, which executes routine volume shifts across plants or markets and adjusts delivery schedules within predefined financial and approval parameters, escalating to human operators when thresholds are exceeded. These task families share characteristics that justify the use of Agentic AI over simpler automation: they require multi-step coordination across fragmented data sources and systems, are repetitive and time-critical, and can operate within explicit safeguards that preserve human authority over high-stakes decisions.

Taken together, the two sets of interviews converge on a single design imperative: the value of Agentic AI in supply chain resilience lies in closing the gap between the speed at which disruptions propagate and the speed at which organizations can detect and respond. The overlap defines the design space: autonomy should be highest where decisions are repetitive, data-intensive, and time-critical, and lowest where stakes are high, or context is ambiguous.

5

First Focus Group

This chapter reports the findings from the first focus group iteration of the study, conducted within the first cycle of the double diamond methodology. Drawing on distinctions between Agentic AI and GenAI identified through expert interviews, particularly in terms of capability, risk, and governability, this chapter examines points of disruption, recovery bottlenecks, and implementation constraints within supply chains. Although the interviews yielded in-depth, domain-specific knowledge, translating these insights into actionable agent design concepts necessitated a structured co-design methodology.

This chapter reports the outcomes of an exploratory focus group conducted within the double-diamond workflow. The focus group fostered shared understanding across SCOR processes by engaging three experts - two with supply chain experience and one with AI expertise - to establish priorities, critically examine assumptions, and evaluate trade-offs in realistic disruption scenarios. In accordance with established best practices for exploratory research, the moderator facilitated active interaction using concrete scenarios. Interactive boards supported online collaboration, enabling participants to build on each other's contributions, propose mitigation plans, and assess feasibility.

The findings provide a set of generalizable design patterns that identify points where information flow deteriorates, decision routines become vulnerable under pressure, and forms of Agentic AI support that are both operationally effective and practically governable. This research makes two key contributions: it identifies four generalizable supply chain resilience failure patterns applicable across diverse organizational contexts, and it introduces three agent design archetypes that align disruption patterns with technical capabilities, each with clearly defined boundaries.

This chapter is structured as follows. Section 5.1 presents the design, execution, and findings of the focus group. Section 5.2 discusses the findings of the follow-up interview, which explores a detailed real-world implementation of an Agent. Section 5.3 provides the chapter's conclusions.

5.1. First Focus Group

Identification of Problems in the Supply Chain

Supply chain fragility arises primarily from coordination failures rather than resource scarcity, such as limited capacity, inventory, or backup suppliers. Although relevant information is available, it is not integrated rapidly enough. Existing procedures often fail to address exceptions, and while risks are detectable, they are not subject to continuous monitoring. Furthermore, lessons learned are not effectively disseminated to prevent recurrence. Instead of replacing human decision-making, agents should enhance coordination by bridging information silos, providing access to exception management protocols, enabling continuous risk monitoring, and minimizing feedback loop delays.

Table 5.1: Supply Chain Resilience Failure Patterns and Agentic AI Design Responses

Element	Synthesis
Pattern 1: Information Fragmentation Under Load	**Problem:** Critical decision-making data is dispersed among disconnected systems, databases, and communication channels. During routine operations, individuals manually bridge these gaps. However, under stressful conditions, the increased coordination load exceeds the capacity for manual integration, resulting in information bottlenecks. **Where it appears:** *Deliver* (transport systems, hub databases, customer records, supplier communications in different formats); *Enable* (dashboard proliferation across procurement, production, logistics, quality); *Return* (claims require correlating defect reports, master data, contracts, financial systems). **Impact:** This approach delays detection, impedes evidence collection for root-cause analysis, and increases recovery time by necessitating sequential rather than parallel problem-solving. **Agentic AI solution:** Agents facilitate the integration of disparate systems by automatically retrieving, correlating, and structuring data, resulting in unified and actionable information views.
Pattern 2: Procedural Rigidity in Exception Scenarios	**Problem:** Standard operating procedures are designed for routine conditions. In the presence of exception scenarios that fall outside established protocols, teams often resort to improvisation, which can result in inconsistent responses and confusion in coordination. **Where it appears:** *Plan* (rare events such as supplier bankruptcy, regulatory change, force majeure lack clear escalation paths/decision authority); *Source* (procurement assumes forecast reliability; large deviations lack procedures for buffer adjustments); *Make* (capacity management handles gradual change but lacks playbooks for sudden bottlenecks). **Impact:** Response latency increases, decision paralysis may occur when authority is unclear, and organizational learning is weakened if ad-hoc actions are not systematically documented. **Agentic AI solution:** Agents provide exception playbooks and institutional knowledge at the point of need, thereby ensuring that infrequent procedures are accessible, standardized, and repeatable.

Continued on next page.

Table 5.1 continued.

Element	Synthesis
Pattern 3: Detection Lag in Slow-Moving Risks	**Problem:** Certain risks develop incrementally over time rather than manifesting as isolated incidents. Periodic manual monitoring typically identifies these risks only after critical thresholds have been surpassed. **Where it appears:** *Make (Waste)* (expiry-driven waste accelerates between weekly samples); *Make (Capacity)* (utilization trends to bottlenecks over days, missed by periodic reviews); *Source* (forecast deviations compound, but monthly reviews delay detection). **Impact:** Delays in response time can lead to increased intervention costs as risks accumulate and can also limit the range of feasible mitigation options. **Agentic AI solution:** Agents facilitate ongoing monitoring and provide early warnings for gradually accumulating risks by identifying deviations prior to the onset of crisis thresholds.
Pattern 4: Feedback Loop Atrophy	**Problem:** Organizational learning relies on effective feedback from operational activities to planning processes. However, these mechanisms often deteriorate because of reporting friction, inadequate system integration, and insufficient incentives for thorough documentation. **Where it appears:** *Return* (manual investigations are time-consuming, causing under-reporting and incomplete attribution); *Enable* (supplier issues logged but not fed back into procurement decisions); *Plan* (reviews occur but do not translate into updated playbooks or process changes). **Impact:** This issue results in repeated failures due to persistent root causes, hinders the refinement of buffers, and diminishes the effectiveness of supplier performance management over time. **Agentic AI solution:** Agents minimize feedback friction by automating documentation processes and integrating operational evidence with planning systems, thereby ensuring that insights are incorporated into revised parameters and playbooks.

Agentic AI Solutions: Mapping Technology to Organizational Fragilities

After identifying organizational weaknesses, the focus group's second phase analyzed the strategic deployment of Agentic AI to address four documented failure patterns: Information Fragmentation Under Load, Procedural Rigidity in Exception Scenarios, Detection Lag in Slow-Moving Risks, and Feedback Loop Atrophy. The resulting Agent concepts are sociotechnical systems designed to augment, rather than supplant, human decision-making.

All three participants independently described Agents as "resilience amplifiers" rather than autonomous decision-makers, viewing them as extensions of organizational sensing and coordination. Participants likened Agents to a supply chain "nervous system" that monitors signals, aggregates fragmented information, and alerts humans when intervention is necessary, while reserving strategic decisions for human oversight.

The Agent concepts, detailed in App. B, were categorized into three groups, each addressing a spe-

cific dimension of resilience. Diagnostic and Monitoring Agents enable early risk detection through continuous surveillance. Coordination and Evidence-Building Agents accelerate disruption recovery by automatically retrieving and synthesizing information from disconnected systems. Interface and Learning Agents disseminate expert knowledge and strengthen feedback loops, ensuring that lessons from disruptions inform future planning.

Table 5.2: Agent Archetypes and Intended Role in Resilience Support

Agent Category	Purpose and Included Agents
1. Diagnostic & Monitoring Agents	**Purpose:** Early warning systems are integrated into operational workflows to mitigate *Detection Lag in Slow-Moving Risks*. This category includes agents that continuously monitor operational metrics and notify stakeholders when thresholds are exceeded or anomalies are detected. Automating routine surveillance of capacity, demand-supply synchronization, and perishability risks enables these agents to prevent gradual progression toward operational constraints that periodic manual reviews may not identify in time for effective intervention. **Included agents:** Capacity–Risk Assessment Agent; Forecast–Monitoring Agent; Waste–Monitoring Agent.
2. Coordination & Evidence-Building Agents	**Purpose:** Recovery from disruptions can be accelerated by retrieving and structuring fragmented information across disconnected systems, thereby directly addressing *Information Fragmentation Under Load*. This category includes agents that aggregate data from disparate sources to generate unified, actionable reports during disruptions. By eliminating manual cross-system data retrieval and automatically compiling comprehensive documentation, these agents reduce coordination effort and reporting friction, both of which typically delay recovery responses. **Included agents:** Documentation and Reporting Agent; Claims Agent; Insights Agent.
3. Interface & Learning Agents	**Purpose:** Make tacit knowledge accessible and strengthen feedback loops to address both *Procedural Rigidity in Exception Scenarios* and *Feedback Loop Atrophy*. This category includes agents that capture expert knowledge, facilitate customer interactions, and enhance the quality of demand signals. By democratizing specialized procedural knowledge and systematically integrating feedback into planning cycles, these agents reduce dependence on specific individuals and ensure that post-event learning is retained rather than lost to underreporting. **Included agents:** Expert–Process Guidance Agent; Feedback Agent; Demand–Forecast Support Agent.

Agent Governance Architecture

Beyond identifying specific Agent use cases, the interviews revealed strong consensus regarding governance requirements for responsible Agentic AI deployment. All three participants independently emphasized that successful implementation requires explicit operational boundaries and multi-layered oversight mechanisms that balance automation benefits against operational, financial, and security risks.

Table 5.3: Read–Act Governance Boundary for Agentic AI

Boundary	Definition and Illustrative Permissions/Prohibitions
Always Allowed	Agents are permitted to perform activities that increase organizational visibility without changing operational state. This includes accessing internal operational data across systems (e.g., production metrics, planning information, historical records, machine status), including confidential information when appropriate technical access controls are in place. Agents may aggregate, reformat, and synthesize dispersed information into unified views to reduce cross-system fragmentation. They may use advanced retrieval methods (e.g., SQL queries, Retrieval-Augmented Generation) to access information beyond standard dashboards. Agents may interact with internal stakeholders and customers to clarify needs, ask questions, and gather context, thereby reducing ambiguity in requirements. They may store relevant information to maintain context across interactions and generate proposals or recommendations for human decision-makers.
Never Allowed	Agents are prohibited from any activity that changes organizational state or commits the organization externally without human authorization. This includes unilateral modifications to operational plans, production schedules, delivery routes, or resource allocations, regardless of Agent confidence. Agents cannot communicate with external parties (e.g., suppliers, partners, customers) without explicit consent, to prevent unauthorized disclosure and preserve human control of external relationships. Financial activities are categorically disallowed: Agents cannot execute transactions, access banking systems, or commit the organization to financial obligations. Agents cannot independently modify master data or share information beyond defined scope, maintaining data integrity and access boundaries. Agents also cannot access supplier systems or proprietary partner data without explicit, scoped permission that respects supplier autonomy. The unifying rule is that consequential actions affecting resources, operations, or external relationships must remain under human control.
Rationale	This governance model maintains human oversight of critical decisions while leveraging artificial intelligence (AI) for its strengths in information processing. AI agents are particularly effective at monitoring diverse sources, identifying patterns within large datasets, and sustaining continuous attention. Nevertheless, decisions involving conflicting priorities, incomplete information, accountability, or interpersonal relationships necessitate human judgment. The read-act distinction ensures that AI agents enhance human capabilities without supplanting human decision-making, thereby providing the advantages of automation while safeguarding human authority.

Oversight and Fail-Safe Mechanisms

The interviews revealed sophisticated thinking about multi-layered oversight systems to ensure AI errors can be detected and corrected before causing cascading failures. The arguments of participants can be classified in three categories of fail-safe mechanisms.

Table 5.4: Oversight Mechanisms for Governable Agent Operation

Mechanism	Operational Requirement and Implementation Logic
Human Validation	When Agent outputs appear incorrect, human operators validate results against current operational reality and verify alignment with authoritative data sources. Review includes checking for systematic issues (e.g., data format errors) and escalating persistent inconsistencies to appropriate management or technical teams. Documentation that may contain confidential information requires human review prior to distribution. In customer-facing applications, complex or ambiguous cases must be transferred immediately to human support via clear and accessible escalation pathways.
Secondary Agent Supervision	A layered control architecture can employ secondary *"control"* Agents to monitor primary Agent behavior while preserving ultimate human authority. Supervisory Agents detect infinite loops or repeated problematic actions, verify that both an AI proposal and explicit human approval exist before any critical change is executed, monitor data-source access patterns for policy compliance, and automatically escalate to humans when predefined thresholds are exceeded. This uses Agentic AI to supervise Agentic AI, but keeps decision authority and accountability with human operators.
Data Quality Controls	To reduce errors driven by poor input data, governance requires mechanisms that detect and correct data quality issues such as misformatted dates, swapped identifiers, outlier values, and stale data. Proposed solutions include anomaly-detection Agents, verified reference databases used for validation, and automatic workflow restarts when data errors are suspected. Persistent or recurring issues must escalate to IT or data management teams to address root causes rather than repeatedly handling symptoms at the operational level.
Escalation Thresholds	Escalation rules define clear conditions for human notification: recommendations that contradict established parameters, unresolvable data quality issues, interactions requiring judgment beyond Agent scope, recurring errors indicating systematic problems, and any situation with financial implications or compliance risk. Escalation targets are context-dependent: operational issues route to production managers, technical failures to IT teams, customer service cases to supervisors, compliance concerns to legal teams, and strategic decisions to senior management.

5.2. A Case Study in Agentic AI for Claims Handling

This case study analyzes an implemented agent-supported claims workflow, demonstrating that Agentic AI design arises primarily from operational constraints rather than from abstract Agentic AI capabilities. The information relevant to this case study was obtained from a semi-structured interview, available in Appendix B, and analyzed using the same methodology applied to the other interviews in this research. Additionally, the analysis was supported by supplementary documents provided by the interviewee. The previous process introduced significant reporting friction, as operators encountered additional procedural steps, unclear requirements, and inadequate time to provide comprehensive ex-

planations. Consequently, defect registration was frequently incomplete. This led to under-reporting of waste, diminished visibility into inventory quality, and distorted supplier performance signals. Over time, fragmented communication and limited responsiveness further undermined reporting effectiveness, as hubs reduced their efforts when tickets did not result in effective follow-up.

A subsequent process redesign enhanced usability by standardizing ticket creation. However, this introduced a trade-off in data quality. While structured categories expedited data entry, they also reduced the amount of free-text context necessary for accurate attribution. Simultaneously, higher ticket volumes and more stringent supplier deadlines rendered manual, case-by-case handling impractical. Automation became essential, not to improve the quality of communication, but to satisfy requirements for speed and scalability while maintaining accountability.

Baseline Workflow: Deterministic Batch Processing

The baseline workflow functions as a scheduled batch process (left workflow). Data are collected from spreadsheets, records are enriched with supplier information through lookups, and inputs are consolidated into standardized records. Routing is determined by hard-coded rules: a central switch matches supplier names and directs items into supplier-specific branches for aggregation. Items that do not match follow a fallback path determined by keyword detection.

While this approach offers predictability, it is constrained when cases are ambiguous, novel, or only partially documented. These cases are assigned to generic categories or require manual correction. Interview findings suggest that this process introduces operational friction. Unclear requirements discourage comprehensive documentation, leading to under-reported waste and weak connections between operations and supplier performance tracking.

Reasons for Redesign

The baseline system did not experience technical failure; rather, it introduced operational friction in areas where reporting processes should have been straightforward. Increased ticket volumes, combined with unyielding supplier deadlines, intensified this issue by widening the gap between manual processing capacity and the required operational speed. Furthermore, expanding the rule tree proved both costly and unstable, as each additional supplier or atypical case increased complexity, maintenance demands, and the likelihood of routing errors.

A further challenge involved data quality. While standardization expedites ticket creation, it diminishes the contextual detail necessary for accurate attribution. Consequently, the redesigned system must function effectively with standardized inputs, explicitly manage uncertainty, and escalate ambiguous cases for human review.

Agentic Workflow: Semantic Classification and Orchestrated Preparation

The agentic workflow transitions from spreadsheet-based batching to ticket-centered orchestration (right workflow). It retrieves records from a ticketing database and processes each record through a comprehensive claim-preparation pipeline. A significant modification is the shift in processing granularity: claims are now managed at the item level rather than as aggregated records. The system decomposes complex SKU fields into individual items, incorporates supplier data, and assigns unique identifiers. This approach enhances traceability and prevents a single ambiguous element from affecting the classification of an entire ticket. Semantic interpretation supersedes fixed routing rules. The

AI agent analyzes operator comments in relation to predefined category definitions, thereby reducing dependence on ever-expanding keyword lists. Generative AI composes claim emails using controlled prompts and structured operational data, which increases both speed and consistency. Accountability decisions are governed by routing logic, escalation protocols, and human oversight, rather than being assigned to unconstrained generative processes. The system automatically manages supporting evidence by collecting images referenced in tickets and compiling email drafts that are ready to send. Human accountability is maintained through explicit escalation procedures. When classification confidence is insufficient, inputs are incomplete, or errors are detected, items are directed to human reviewers. This transition shifts operator responsibilities from repetitive tasks to supervision, governance, and decision-making in ambiguous situations.

Grounding with Structured Data

The agent accesses operational data directly through database connections, retrieving internal ticket fields and master-data attributes as structured context for routing and email generation. This grounding enhances operational accuracy by ensuring the use of authoritative identifiers, maintaining explicit links to original tickets, and minimizing fabrication risk by restricting text generation to verified variables.

Comparison to Ungrounded LLM Use

In the absence of grounding, a general-purpose LLM may generate fluent text but does not have assured access to internal identifiers, supplier master data, or policy rules. This limitation increases the risk of operationally incorrect outputs. Structured grounding enhances factual accuracy and processing efficiency, thereby reducing the disparity between well-written language and accurate claims handling.

Operational Impact and Supply Chain Resilience

The Agentic AI workflow organizes incoming tickets by segmenting them at the item level and categorizing cases according to problem type and responsible party. It then prepares near-complete claims by compiling relevant evidence and drafting communications to suppliers. Human oversight remains integral through explicit escalation protocols and final review of ambiguous or high-risk cases, shifting responsibilities from repetitive processing to supervisory governance and exception management. This approach positions Agentic AI as an orchestration and routing solution, grounded in operational data and constrained by escalation and control mechanisms. Generative models are employed in a limited, supportive capacity to draft outputs, rather than to make core accountability decisions. This implementation enhances supply chain resilience through multiple mechanisms. Accelerated and standardized claims processing reduces the time between defect detection and supplier accountability, thereby increasing organizational responsiveness to disruptions. Comprehensive documentation and accurate routing improve visibility into supplier performance trends and recurring issues, supporting more informed decisions regarding supplier relationships and risk mitigation strategies. Transitioning from under-reporting to comprehensive documentation provides a more accurate assessment of operational waste and its origins, facilitating proactive supplier management instead of reactive problem-solving. Shortened intervals from issue detection to claim submission enable the organization to meet contractual deadlines, safeguard financial recovery, and maintain stronger negotiating positions. Automated evidence compilation and enhanced traceability further reinforce the organization's stance in disputes and promote continuous improvement by revealing root cause patterns throughout the supply chain. The escalation mechanism ensures that cases involving accountability concerns receive appropriate human oversight, thereby maintaining supplier relationships while pursuing legitimate claims. This balance between automation and human judgment preserves supply chain partnerships and enables

systematic resolution of performance issues.

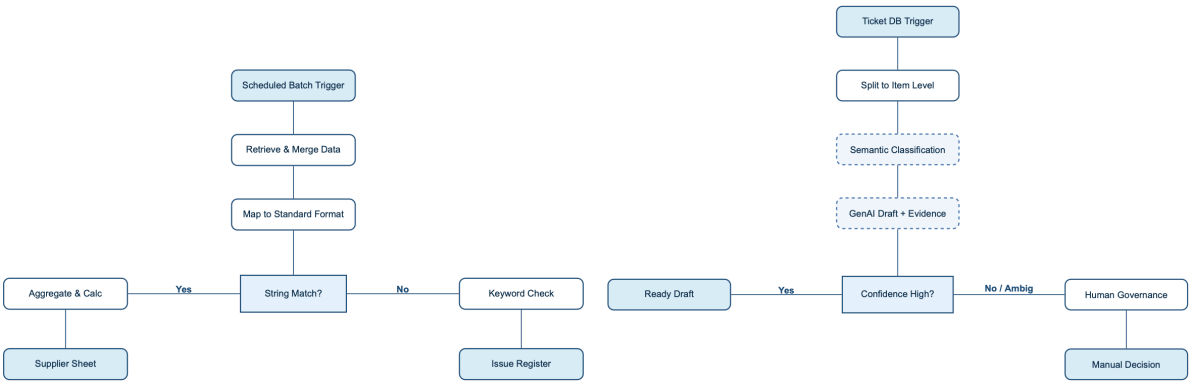


Figure 5.1: Visual comparison of Baseline Workflow and Agentic Workflow

5.3. Concluding Remarks

This chapter translates conceptual possibilities into actionable designs. The focus group required abstract concepts to address real-world disruption scenarios, resulting in agent concepts grounded in operational breakdowns rather than being restricted to technological capabilities.

The four identified failure patterns demonstrate that coordination, rather than resource availability, is the primary factor affecting supply chain resilience. Information becomes fragmented across systems when manual integration is insufficient. Standard procedures are ineffective when exceptions surpass established protocols. Risks accumulate undetected between periodic reviews. Operational insights often fail to reach planning systems due to reporting barriers. Increasing inventory or adding backup suppliers does not resolve these underlying issues.

The three agent archetypes directly address these challenges. Diagnostic agents enhance sensing capacity through continuous monitoring. Coordination agents reduce recovery time by consolidating dispersed information. Interface agents increase access to expert knowledge and reduce documentation requirements. Their implementability is supported by governance mechanisms developed during the design process. The read-act boundary serves as a technical specification, enabling agents to access data and generate recommendations while preventing them from altering operational states or allocating resources without human approval. Multi-layered oversight introduces redundancy, mitigating the impact of inevitable imperfections in Agentic AI outputs and input data. The claims-processing Agent serves as a real-world implementation example.

The claims-processing Agent case study demonstrates the practical effectiveness of this approach. Rather than relying on rigid if–then rules, the Agent employs classifications based on semantic meaning, enabling it to accommodate variations in how issues are described. Each claim is treated as an individual case, preventing unclear or atypical examples from influencing other decisions. The Agent generates outputs solely from verifiable data, thereby reducing hallucinations and unsupported conclusions. When a case is ambiguous or requires trade-offs, the system escalates it to a human operator. These design choices were implemented to address real operational challenges, such as underreporting, missing context in tickets, and excessive case volumes for manual processing, rather than to highlight Agentic AI capabilities.

6

Second Iteration Interviews - AI, Supply Chain and Avionics Experts

This chapter reports the findings from the second interview iteration of the study, conducted within the second cycle of the double diamond methodology. The primary objective is to validate the proposed agents in practical contexts by identifying areas where they can substantially enhance supply chain resilience and where their impact may be limited. Unlike the initial qualitative round, which focused on idea generation, this phase emphasizes validation and refinement.

The supply chain expert interviews serve three primary purposes: (i) to assess whether the proposed agents plausibly enhance resilience across the disruption patterns identified by Ivanov and Sokolov (2010), (ii) to clarify each agent's strengths and weaknesses, and (iii) to collect specific recommendations regarding system integration, safeguards, and design improvements.

The AI expert interviews address conceptual and technical aspects that were not fully explored in the initial round. The objective is to refine the rationale for technology selection and to define deployment conditions. Specifically, these interviews investigate the distinction between Generative AI and Agentic AI, thresholds for autonomy, governance structures, safeguards, appropriate modes of human involvement, explainability requirements for different stakeholder groups, organizational and infrastructure readiness, trade-offs between performance and speed including human-factor risks, and the maturity and reliability challenges associated with multi-layer and multi-agent systems.

This iteration also includes interviews with Airbus professionals and a technical analysis of the Airbus Automatic Flight System to examine the operation of mature automation architectures in high-risk, safety-critical environments. This cross-industry perspective, based on the analogical reasoning methodology described in Section 3.5, establishes a governance benchmark for applying principles from aviation to supply chain resilience applications.

6.1. Second Iteration - Supply Chain Experts

The findings from the second iteration of supply chain expert interviews confirm previous conclusions and introduce significant extensions, advancing the discussion from diagnostic framing to integration, anticipation, and action. The primary contribution is the identification of an architectural gap between monitoring and response. Current agent designs are effective at detection but lack mechanisms for translating alerts into actionable responses. Both experts emphasized that agents should not only monitor but also execute bounded responses. They proposed a two-tier architecture in which monitoring agents detect and alert, while action agents consume those alerts to trigger re-routing, re-allocation, or maintenance interventions. Without this action layer, organizations risk alert fatigue as monitoring capabilities expand without corresponding response mechanisms.

A closely related theme is the imperative for integration, which emerged from both interviews. The experts argued that isolated agents create new silos and that value is realized only when agents operate collaboratively rather than independently. The first expert consistently advocated for linking agents to related functions, such as connecting demand forecasting to stock management and purchasing, and enabling monitoring agents to trigger guidance agents based on events rather than requests. The second expert supported the development of multi-signal agents capable of jointly reasoning about supply, demand, and visibility data. These insights suggest that unchecked agent proliferation without integration may increase coordination burdens rather than alleviate them. Design priorities should therefore favour a smaller number of well-integrated agents over numerous specialized but disconnected ones.

The second iteration introduces an explicit temporal dimension that was absent from earlier phases. The second expert recommended incorporating forward-looking simulation components to project how situations may evolve over the coming days. The first expert suggested enhancing demand models with signals beyond historical data, while also acknowledging the limitations of assuming future conditions will mirror the past. This approach shifts agent design from retrospective analysis and present-focused monitoring toward genuine anticipation. It enables the detection of risk trajectories before thresholds are breached through scenario-based forecasting and pattern recognition across multiple converging signals. In addition to this temporal extension, both experts recognized that generic agent patterns offer limited value. They advocated for subdomain-specific implementations in areas such as quality management, process analysis, and supplier performance, rather than relying on generic documentation agents. This perspective supports a platform-plus-modules architecture, featuring a common infrastructure with pluggable domain-specific capabilities.

Regarding human oversight, the second iteration refines earlier frameworks by introducing important calibration. The first expert stated that human controls should be removed only when automation can be demonstrated to be fail-safe, citing pharmaceuticals as an example where permanent oversight is required regardless of system maturity. The second expert emphasized the necessity of quality checks and systematic human review during initial adoption, noting that oversight may be reduced as systems mature but should never be eliminated entirely. This positions human-in-the-loop oversight as a spectrum, calibrated according to safety criticality, financial impact, maturity stage, and decision reversibility, rather than as a binary choice. Additionally, the second iteration extends feedback mechanisms toward root-cause correction. The second expert proposed that human corrections should be integrated into master data management to address issues at their source, ensuring that operational learnings are translated into parameter updates to prevent recurrence.

A novel contribution of this iteration is the extension to cross-organizational scope. The first expert proposed agents that operate beyond individual firm boundaries, specifically suggesting an agent that continuously monitors regulations and external constraints across regions to assist companies in reconfiguring supply chains in compliance with these requirements. Such an agent would function at a higher level than individual firms and would be operated by an entity not in direct competition, such as an industry consortium or a neutral third party. Additionally, the second iteration broadens governance considerations from technical architecture to organizational processes. Experts called for strict rules regarding acceptance criteria and escalation triggers, harmonization across teams, formal sign-off from relevant parties, and the integration of agents into organisation-wide processes rather than treating them as isolated technical deployments.

6.2. Second Iteration - AI Experts

This section reports findings from interviews with AI experts conducted during the second iteration of the double diamond methodology. The interviews validate and expand upon foundational insights from the first iteration, specifically regarding Agentic AI definitions, failure modes and governance requirements. In addition to confirming core principles, the second iteration identifies new dimensions, including task complexity as a criterion for technology selection, operational infrastructure requirements, differentiated oversight configurations, and enhanced robustness through simulation and confidence signals.

Confirmation of First Iteration Findings

The second iteration reinforces three central themes identified in the first iteration. First, in the context of the GenAI-Agentic AI relationship, participants confirmed that Agentic AI serves as an operational layer built upon Generative AI. Agentic AI is distinguished by its ability to perform multi-step execution, rather than solely generating content. The primary distinction is based on functional requirements: whether the system is required only to generate content or to reliably execute task steps as well. Second, the principle of risk-based autonomy was explicitly operationalized. Participants recommended establishing clear thresholds to differentiate low-impact actions, which can proceed without escalation, from high-impact decisions that require human review. This approach results in the implementation of guardrails and structured exception flows that direct high-risk cases to human reviewers. Third, concerns regarding failure modes and error propagation were reaffirmed. Participants noted that even GenAI, which is less complex than Agentic AI, already demonstrates numerous failures. This observation underscores the need for high caution in more complex deployments, where a single error may lead to cascading failures.

New Dimensions from the Second Iteration

Building on previous findings, the second iteration introduces six interrelated deployment dimensions that operationalize the governance architecture. Task complexity serves as the primary criterion for technology selection. For simple prediction tasks, such as ranking, scoring, or prioritizing, standard machine learning is generally sufficient. Generative AI is suitable for bounded, single-step interactions where errors are easily identified and corrected. Agentic AI is justified only for tasks involving multi-step, multi-level reasoning that require significant coordination and decision logic. The guiding principle is to select the simplest approach that fulfills the requirements, as unnecessary complexity increases costs, integration challenges, and failure risks without delivering proportional benefits. This selection logic is effective only when supported by appropriate operational infrastructure. Evidence highlights three

essential prerequisites: high-quality, reliable data; robust system interfaces that prevent fragile point-to-point integrations; and dedicated testing environments for safe verification prior to deployment. In practice, enterprise readiness encompasses not only secure data handling and model performance, but also the integration of security with deep connectivity. Business value is realized when models are embedded within internal tools and core platforms, such as ERP systems, rather than remaining as isolated prototypes. Within this infrastructure, the second iteration distinguishes between continuous approval and scalable control in human oversight. Human-in-the-loop models necessitate approval for every decision, whereas human supervision involves periodic monitoring and targeted intervention in response to risk signals, anomalies, or boundary conditions. For Agentic AI, supervision is preferable, as requiring approval at every step would negate the efficiency and coordination advantages of automation. Oversight practices also evolve with system maturity: initial deployments require frequent validation, while mature systems can be monitored less intensively, though the option for intervention remains essential. Since oversight involves multiple stakeholders, explainability must be designed for diverse audiences rather than treated as a generic addition. Planning stakeholders require explanations focused on feature-outcome relationships and operational drivers, while data scientists prioritize reliability, performance metrics, and failure modes. Both groups consistently need contrastive explanations, clarifying why a particular outcome occurred and why plausible alternatives did not. These explanations should be provided both locally, at individual decision points, and globally, to ensure transparency across the entire workflow. Explanations serve not only as documentation but also as tools to refine decision logic and minimize recurring errors. To prevent improved reasoning from diminishing responsibility, the iteration formalizes accountability and safeguards. Deployments should operate within a clear accountability framework that assigns ownership of outcomes and responsibility for failures, supported by monitoring systems that facilitate incident reconstruction and auditability. Operational continuity relies on predefined fallback options that maintain processes when an agent fails. Consequently, safeguards are positioned as mechanisms for organizational resilience, not solely as technical recovery solutions.

6.3. Architectural Refinements to the Agent Archetypes

The second iteration presents a series of refinements to the agent archetype design that are applicable across all three categories.

The most significant refinement is the governance constraint imposed on the response tier. While Response and Coordination Agents introduce an action-taking capability that was absent from the original design, this capability remains deliberately limited and is subject to rigorous human validation at every critical step. As detailed in the Final Agents Portfolio (Appendix C.5), agents within this archetype are consistently authorized to retrieve, reconcile, simulate, and recommend, but are never permitted to implement high-impact changes, modify financial or master records, or release outputs externally without explicit human approval. Escalation protocols and rollback procedures are predefined for each agent, ensuring that anomalous outputs or threshold breaches are directed to human reviewers rather than resolved autonomously. This constraint is not a temporary measure to be relaxed as systems mature; rather, it represents a principled governance stance confirmed through both supply chain and AI expert interviews. The efficiency gains of the response tier derive from automating preparation and coordination, not from eliminating human judgment in consequential decisions.

Three further refinements apply across all archetypes. First, agents should not operate in isolation: value is realized only when agents share signals and trigger one another in response to events rather

than requests, and unchecked proliferation of disconnected agents risks increasing coordination burdens rather than reducing them. Second, a forward-looking dimension should complement retrospective monitoring: incorporating simulation components and multi-signal pattern recognition shifts agent design from detecting present conditions toward anticipating risk trajectories before thresholds are breached. Third, a platform-plus-modules architecture — featuring shared infrastructure with pluggable domain-specific capabilities — is preferable to fully generic agents, enabling subdomain implementations in areas such as quality management, supplier performance, and process analysis without duplicating governance overhead.

These refinements do not invalidate the original archetype structure; rather, they extend it. The three categories remain architecturally coherent, but their internal design should incorporate action execution, inter-agent collaboration, temporal anticipation, and domain specificity to fully realize their potential in supporting supply chain resilience.

6.4. Temporal Alignment of Agent Archetypes with the Disruption Profile

The focus group and the claims-processing case study identified three agent archetypes. To complement this process-oriented perspective, this section positions the three archetypes on the temporal disruption profile introduced in Chapter 2, analyzing when during a disruption each archetype becomes most critical. Figure 6.1 illustrates this mapping.

Diagnostic and Monitoring Agents, operate primarily in the pre-disruption and early detection phases (phases 1–3). These agents serve as continuous background monitors, tracking capacity utilization trends, forecast deviations, and perishability indicators that may be missed by periodic manual reviews. By identifying gradual risk accumulation before the performance curve declines, they extend the window for preventive intervention and can reduce the severity of the eventual performance drop. Focus group evidence supports this positioning, as participants described scenarios where weekly or monthly review cycles failed to detect risks compounding over several days.

Response and Coordination Agents, are most relevant during the performance trough and recovery phases (phases 4–7). Focus group findings indicate that information fragmentation intensifies when coordination demands peak, as decision-makers require rapid access to data distributed across disconnected systems and manual cross-referencing becomes a bottleneck. The claims-processing case study in Section 5.2 demonstrates this effect: the agentic workflow shortens the interval from defect detection to claim submission, resulting in a steeper recovery slope and a reduced overall performance deficit.

Interface and Learning Agents have their primary contribution occurring in the long-term impact phase (phase 8). At this stage, the organization determines whether it returns to its pre-disruption baseline or progresses to an improved operating state, consistent with the growth trajectory identified by Hohenstein et al. (2015). These agents are critical in connecting phase 8 back to phase 1 by transforming disruption experience into updated playbooks and institutional knowledge, thereby reinforcing resilience as a continuous cycle.

This mapping demonstrates that the archetypes are not interchangeable; each operates within a distinct temporal window where its associated failure pattern is most pronounced. Detection lag accumulates

before disruptions materialize (phases 1–3), information fragmentation intensifies under load during impact and recovery (phases 4–7), and feedback loop atrophy emerges after the crisis, when organizational attention shifts before lessons are captured (phase 8).

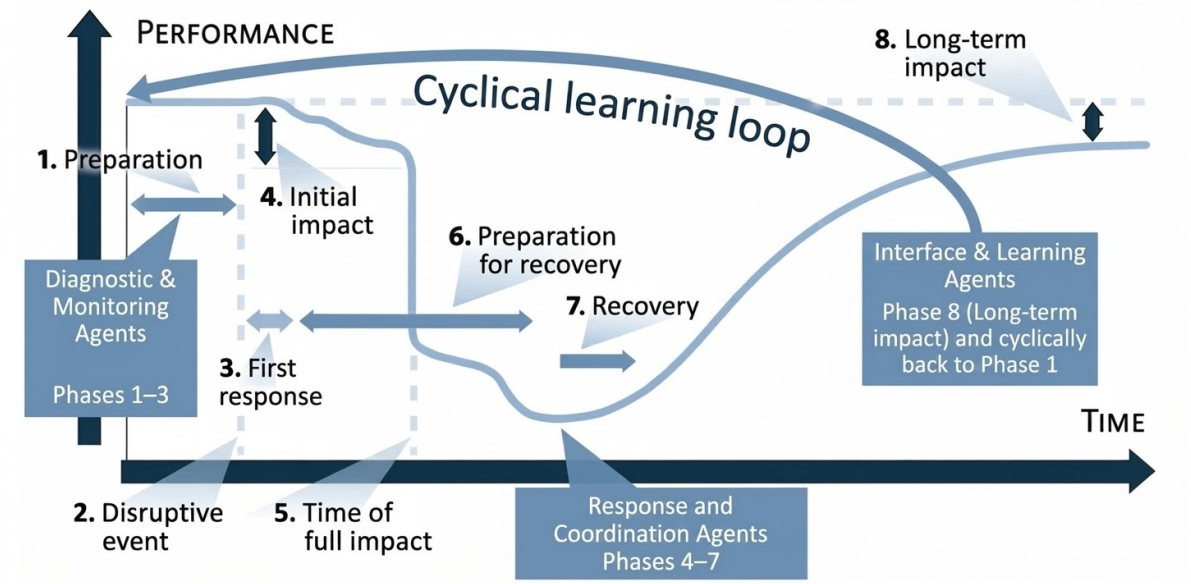


Figure 6.1: Agent archetypes mapped onto the disruption profile phases. Disruption profile concept based on Sheffi and Rice (2005).

6.5. Recovery Trajectory Exceptions and the Role of Agent Archetypes

While the disruption profile presented in Chapter 2 provides a stylised representation of supply chain performance over time, the resilience literature identifies several exceptions to a smooth recovery path. As noted by Hohenstein et al. (2015) and Hendricks and Singhal (2005), post-disruption recovery may follow fundamentally different trajectories, including recovery to an improved baseline or persistent performance loss. This section examines three such exceptions and evaluates which agent archetypes identified in this study offer direct support for addressing them.

The Resilience Triangle

The resilience triangle measures cumulative performance loss as an area defined by the depth of the performance drop and the length of the recovery window. The three archetypes identified in this study each act on a distinct dimension of this area. Diagnostic and Monitoring Agents compress the triangle vertically: by detecting risk accumulation before it translates into a full performance drop, they reduce how far the performance curve falls. Response and Coordination Agents compress it horizontally: by reducing the time between disruption onset and effective operational response, they steepen the recovery slope and shorten the period before normal performance is restored. Interface and Learning Agents act prospectively rather than on the current triangle: by retaining and institutionalizing lessons from each disruption, they reduce the baseline vulnerability that determines how large future triangles will be when the next disruption occurs.

The "Bounce Forward" Trajectory

Rather than a simple restoration to the prior baseline, recovery can follow a “bounce forward” trajectory, consistent with the definition of resilience as a system’s ability to not only return to its original state but to adapt, grow, and move to a new, more desirable state after being disturbed (Hohenstein et al., 2015). Interface and Learning Agents are the archetype most directly aligned with enabling this outcome. By systematically translating disruption experience into updated playbooks and institutional knowledge, they close the feedback loop between phase 8 and phase 1 of the disruption profile. Without this archetype, organizations risk returning only to their prior baseline, missing the opportunity for structural resilience improvement.

After-Shock Risks

In an evolving crisis, resuming operations too quickly can lead to renewed disruptions, as rushed production resumptions without adequate risk mitigation can trigger new disruption clusters and subsequent operational shutdowns (T. Fan et al., 2023; Ivanov, 2021a). Diagnostic and Monitoring Agents are the archetype best positioned to mitigate this risk. Their continuous monitoring of capacity utilization and operational indicators during the recovery phase can flag premature resumption signals, providing decision-makers with objective, real-time evidence before restart decisions are made - a function that periodic human review cycles are structurally unable to fulfil.

Disruption Tails

Supply chains face the risk of “disruption tails,” defined as postponed effects or residues from the initial disruption period, such as backlogs and delayed orders (Ivanov, 2021a). If management policies are not adapted accordingly, these tails can severely destabilise inventory and production dynamics during the recovery phase (Ivanov, 2021a). Response and Coordination Agents are the archetype most directly equipped to address this exception. Their capacity to aggregate and cross-reference data across disconnected systems maintains ongoing visibility over backlog accumulation and delayed orders that would otherwise remain invisible until they cascade into secondary disruptions.

6.6. Automatic Flight System Architecture and Safety Design: A Cross-Industry Analogical Reasoning Perspective

In addition to the technical description of Airbus automation, which can be found in Appendix C.6, this chapter incorporates two semi-structured interviews that illustrate how the system is experienced in operational contexts and justified within engineering practice. These interviews serve as qualitative evidence to identify mechanisms that render high automation governable in safety-critical environments and to distill these mechanisms into transferable principles for supervised human–machine decision support.

The first interview, conducted with an Airbus line pilot, focuses on the operational use of automation. It details how flight crews monitor and prioritize anomalies using the Electronic Centralized Aircraft Monitor (ECAM), manage mode awareness and workload, and implement standard procedures, cross-checks, and tactical simplification when aircraft behavior diverges from expectations. The second interview, conducted with an Airbus AI and automation engineer, elucidates the design constraints influencing these operational behaviors, such as certification requirements, conservative integration of learn-

ing systems, redundancy and fault isolation logic, and post-event analysis through logs and traceable feedback mechanisms. Collectively, these perspectives present autonomy as structured, bounded, and recoverable, rather than open-ended, and inform the themes and design patterns discussed in the remainder of the chapter.

Table 6.1: Themes and Definitions – Airbus Pilot Interview

Theme Name	Definition
Theme 1: ECAM-centred monitoring and graded alerting	A central ECAM-based monitoring system initially detects, classifies, and prioritizes abnormal situations. This system codifies aircraft behavior into explicit, ordered messages and employs graded visual and alerts to direct pilot attention to the most safety-critical failures while preventing information overload.
Theme 2: Automation by default and layered control authority	Normal operation relies on extensive automation, incorporating layered protections and redundancy to maintain trajectory and stability. Autopilot and flight-computer systems are activated early, direct pilot actions according to established priorities, and are temporarily disengaged or degraded if operational limits are exceeded or data become unreliable. These systems are subsequently re-engaged once stable conditions are restored.
Theme 3: Tactical simplification via knobs, semi-auto modes and reduced typing	In situations characterized by time constraints or increased task demands, crews temporarily simplify control strategies by employing semi-automatic modes, utilizing knob-based selections, and dividing labor. These approaches help avoid slow keypad entries and maintain task sequence, while also allowing for manual control of brief segments when automation continues to provide operational benefits.
Theme 4: Proceduralisation, briefing and governance of responses	Responses to abnormal situations are highly proceduralized and collectively managed. Actions adhere to predefined memory items and checklists, are practiced in simulators, incorporated into comprehensive briefings and diversion planning, and are continuously improved through a feedback loop between the original equipment manufacturer and operators, rather than relying on individual pilot improvisation.
Theme 5: Cross-checking, plausibility checks and manual verification of automation	Automation outputs and system indications are continuously cross-checked with pilot expectations, raw parameters, and relevant documentation. When guidance appears suspect, procedures include verifying data-entry accuracy, conducting dual-crew verification, monitoring essential engine and attitude indicators, observing color changes as early warnings, and consulting system manuals with targeted chapter references as necessary.

Table 6.2: Themes and Definitions – AI and Automation Engineer Airbus

Theme Name	Definition
Theme 1: Conservative use of learning systems in critical control	Safety-critical aircraft functions continue to operate within a traditional, deterministic, and certifiable software framework. Learning systems are permitted only if their behavior can be exhaustively tested across the entire input space or if their failure probability is demonstrated to be below stringent thresholds; otherwise, such systems are excluded from the control loop.
Theme 2: Dual role of AI: development tool and non-critical adviser	Learning systems serve as developmental support tools and advisory layers within non-critical domains. They assist in generating and optimizing code that is subsequently certified in deterministic form. Additionally, these systems are integrated into peripheral, non-critical functions, such as vision alerts, where they undergo offline retraining cycles and lack direct control authority.
Theme 3: Redundancy and fault isolation as primary safety net	System safety is primarily ensured through architectural redundancy, diversity, and isolation, rather than relying on adaptive or self-healing artificial intelligence. Command and monitor patterns, diverse implementations, and voting logic are employed to detect faults, deactivate malfunctioning units, transition to degraded operational modes, and prevent the propagation of single failures throughout the system.
Theme 4: Human-centred authority and pilot engagement	Automation systems are developed based on the principle that the pilot retains ultimate authority. The system architecture provides multiple manual override options, degraded operational modes, and clear status indications. Additionally, operational procedures and phase-based task allocation, such as manual control during take-off and landing and automation during cruise, are intended to maintain manual flying proficiency and ensure pilots remain prepared to assume control when necessary.
Theme 5: Traceability, learning from failures and specification updates	When anomalies occur, systems generate logs that enable engineers to reconstruct events following the flight. During real-time operations, the emphasis remains on providing clear actions for the crew. Subsequent in-depth analysis is conducted on the ground, frequently identifying specification gaps. These findings inform enhancements to requirements, design, and certification processes, with traceability prioritized for significant and critical issues.

6.7. Concluding Remarks

This chapter presents findings from the second round of expert interviews and a cross-industry analysis of the Airbus Automatic Flight System. It delineates the practical boundaries and design requirements for deploying Agentic AI to strengthen supply chain resilience.

A principal finding is the architectural gap between monitoring and response. Existing agent designs

detect issues but lack mechanisms to translate alerts into concrete actions. The proposed two-tier architecture addresses this gap by pairing monitoring agents with action agents. The Airbus layered control stack exemplifies this separation: the Flight Management System computes targets, Flight Guidance translates these targets into commands, and Fly-By-Wire executes commands while enforcing protections. Effective automation depends on establishing clear boundaries between deliberation, guidance, and execution.

Both expert groups concluded that isolated agents increase coordination burdens rather than reducing them. Value emerges when agents collaborate, share signals, and trigger actions in response to events. Airbus achieves this through independently developed, parallel-operating systems that employ voting logic. Divergent outputs function as fault signals, enabling the system to isolate malfunctioning units before errors propagate. This pattern supports deploying different models or architectures for the same critical task and cross-validating their outputs. Supply chain experts extended this logic beyond individual firms by proposing cross-organizational agents operated by neutral entities, such as industry consortia.

The second iteration introduces a temporal dimension that was absent from earlier phases. Experts recommended forward-looking simulation and multi-signal pattern recognition to shift agent design from retrospective analysis toward anticipation of risk trajectories before thresholds are breached. The Airbus validation methodology reinforces this approach through synthetic scenario generation, digital twins, and hybrid simulation rigs that validate system behavior under conditions impractical or impossible to reproduce in real operations. For supply chain applications, Agentic AI systems should be stress-tested against synthetically generated disruption scenarios, including compound, cascading, and unprecedented events, prior to deployment.

AI experts contributed a complementary selection principle: task complexity should determine technology choice. Standard machine learning suffices for simple predictions, Generative AI is suitable for bounded single-step interactions, and Agentic AI should be reserved for multi-step, multi-level reasoning. The guiding principle is to select the simplest adequate approach, as unnecessary complexity increases costs, integration challenges, and failure risks. This approach is effective only when supported by high-quality data, robust system interfaces, and dedicated testing environments.

The second iteration refines human oversight from a binary choice to a calibrated spectrum. AI experts distinguish between human-in-the-loop arrangements, which require constant approval, and human supervision, which involves periodic monitoring with intervention triggered by risk signals. For Agentic AI, supervision is preferable because constant approvals diminish the efficiency of automation. Nevertheless, oversight should always be maintained. The Airbus control laws implement this principle through graceful degradation: Normal Law provides full protection with bounded autonomy, while Alternate and Direct Laws progressively reduce automation but maintain pilot authority. Explainability should be tailored to diverse stakeholder groups at both individual decision points and the system level. Robustness must be treated as an ongoing process, emphasizing confidence signals for out-of-scope conditions and feedback loops that enable learning from human corrections.

The Airbus case also provides process-level lessons for AI in supply chains. The conservative integration philosophy advocates a staged pathway in which agents begin in advisory roles and gain execution authority only after exhaustive validation. Operationally, responses to abnormal situations are highly proceduralized, collectively managed, practiced in simulators, and continuously improved through manufacturer–operator feedback loops. Automation outputs are cross-checked against pilot

expectations and raw parameters at every phase of flight, establishing proactive verification as standard practice. Under high workload, crews simplify control by switching to semi-automatic modes rather than pursuing full optimization. The engineer interview highlights a temporal separation between real-time response and post-event learning. Clear procedural actions are taken during operations, with in-depth analysis and specification updates conducted after the event. These practices demonstrate that the effectiveness of the eight pillars depends not only on technical architecture but also on organizational routines, proceduralized responses, continuous verification, and structured post-event learning.

In summary, governance challenges for Agentic AI in supply chain resilience are primarily architectural and organizational rather than technical. The eight pillars derived from Airbus practice translate these principles into actionable design patterns. The agents proposed in this thesis constitute an initial step, and their value will depend on integration into organizational processes rather than deployment as isolated technical solutions.

7

Disruption Readiness for Agentic AI: Evidence from a Second Focus Group

This chapter reports the findings from the second focus group iteration of the study, conducted within the second cycle of the double diamond methodology. As outlined in the previous chapter, Airbus employs synthetic data, scenario generation, and digital twin technology to prepare automation systems for unforeseen situations and to meet the rigorous safety standards of high-risk aviation operations. Automation failures in aviation can result in catastrophic outcomes, endangering lives on every flight. Physical flight tests are neither safe nor economically feasible for replicating the full spectrum of rare equipment failures, extreme weather events, or edge-case scenarios that automated systems may encounter over millions of flight hours. Allowing such events to occur naturally during service would subject passengers and crew to unacceptable risk. Consequently, aerospace manufacturers have adopted advanced simulation-based methodologies to systematically explore operational boundaries within virtual environments prior to real-world deployment. This approach, supported by comprehensive synthetic scenario coverage and digital twin validation, enables the industry to certify that automation meets the exceptional reliability standards required for rapid machine decision-making when human lives are at stake.

While supply chain disruptions do not typically present immediate life-or-death consequences, the fundamental challenge of preparing autonomous systems for rare, high-impact events remains critical. Previous chapters have identified key obstacles in deploying Agentic AI to enhance supply chain resilience: these systems must address unexpected disruptions such as supplier bankruptcies, geopolitical shocks, and unprecedented demand volatility. A central question remains: how can Agentic AI systems be prepared for scenarios they have not previously encountered? As in aviation, where waiting for catastrophic failures is unacceptable, supply chains cannot rely on AI agents learning solely from real-world crises that may result in significant financial losses or operational disruptions. Both domains require systems capable of reliable performance in the long tail of improbable yet consequential scenarios.

This chapter examines methods for developing Agentic AI systems that sustain resilience when con-

fronted with previously unseen disruptions. The approach parallels pilot training: although aircraft malfunctions are infrequent, pilots invest substantial time in simulators to practice emergency procedures. Similarly, AI agents require systematic preparation for rare supply chain crises to ensure effective responses. Drawing on insights from a validation focus group with AI and supply chain experts, follow-up interviews with supply chain professionals, and established methodologies from aviation and machine learning research, this chapter synthesizes three principal approaches. First, it analyzes the use of synthetic data to generate realistic scenarios for events that have not yet occurred. Second, it investigates digital twins as safe testing environments where agents can learn without jeopardizing actual operations. Third, it introduces epistemic-aware methods that enable AI systems to recognize novel situations beyond their training. The chapter concludes by integrating practical implementation insights, emphasizing that disruption readiness is a continuous system design challenge requiring detection mechanisms, conservative fallback strategies, and ongoing learning from unexpected events.

7.1. Building Disruption Readiness into Agentic AI: Conclusions from the Validation Focus Group

A validation focus group composed of AI and supply chain experts identified several core principles for designing Agentic AI systems to improve supply chain resilience during disruptive events. Structured discussions, collaborative exercises, and plenary sessions enabled participants to reach a consensus that challenges conventional approaches to preparing AI for rare events.

Disruption Readiness as System Design

Participants indicated that disruption readiness should be conceptualized as the systematic design of the entire socio-technical system that supports a largely fixed foundational model. The group observed that disruptions, which are outliers beyond historical distributions, cannot be reliably detected using traditional statistical training methods. A model lacking exposure to disruption classes cannot accurately delineate 'normal' operational boundaries. Consequently, disruption readiness was characterized as a system-level design challenge that requires controlled inputs, explicit testing protocols, robust governance structures, and structured human-AI interaction.

Through analysis of practical examples and operational constraints, the focus group identified three complementary strategies as essential for embedding disruption readiness in Agentic AI systems. The first strategy involves synthetic scenario creation, where teams deliberately construct extreme or counter-intuitive scenarios, such as unlikely geopolitical events or improbable combinations of shocks, to generate synthetic datasets of disruptive events. These scenarios allow systems to explore responses to crises, such as canal blockages or pandemics, without reliance on real-world occurrences. The objective is not to replicate reality precisely, but to systematically expose systems to a broad spectrum of volatility and stress conditions, thereby clarifying appropriate responses to unexpected events.

Expanding upon the synthetic scenario approach, participants recommended the implementation of end-to-end unit testing for edge cases. This strategy shifts the emphasis from optimizing generic metrics during model development to defining critical edge-case scenarios with predetermined desired outcomes. The group advocated for subjecting agents to these predefined situations to evaluate whether they consistently guide systems toward acceptable solutions. This end-to-end unit testing approach reorients evaluation from statistical performance toward stress-testing system behavior under crisis

conditions.

The focus group identified shadow deployment as a particularly promising strategy, citing examples from Dutch Railways. In this approach, agents operate in parallel with human operators or legacy systems, receiving identical real-time inputs and generating proposed actions without actual execution. This arrangement enables direct comparison between agent and human responses during real disturbances, such as network disruptions, crew shortages, and rolling-stock issues, while mitigating operational risk. However, the group emphasized several caveats. Participants cautioned that synthetic data may not generalize effectively if simulated scenarios do not capture the full physical, geopolitical, or organizational complexity of real disruptions, potentially leading to optimization in an oversimplified context. The group also noted that developing high-fidelity simulations is resource-intensive and introduces a significant sim-to-real gap, which must be explicitly addressed when making resilience claims.

Strategic Design Principles

Participants reached strong consensus that developing a single, generic 'resilience agent' for global supply chains is neither practical nor beneficial, consistent with findings from previous interviews and focus groups. The group argued that such systems would face excessive complexity and produce vague or generic recommendations. Instead, participants emphasized the importance of defining specific, operationally bounded problems, such as reallocating orders among a defined set of alternative suppliers or realigning schedules and capacity after operational disturbances. Within clearly scoped decision spaces, Agentic AI was considered more effective and easier to evaluate, as acceptable actions, constraints, and success criteria can be explicitly defined.

The focus group categorized resilience challenges by type and concluded that not all tasks are equally suited to GenAI and Agentic AI approaches. For anomaly detection, several participants noted that traditional statistical methods and classical machine learning models often outperform GenAI in accuracy and robustness. However, once disruptions are detected and characterized, Agentic AI was considered valuable for orchestrating the multi-step decisions required to mitigate impact, such as re-routing flows, adjusting capacity, prioritizing customers, and coordinating information among stakeholders. This distinction reinforced the conclusion that GenAI and Agentic AI should be deployed where they offer comparative advantages, rather than serving as the default choice for all resilience tasks.

For demand-related challenges, participants with AI expertise argued that LLM-based models are not optimal for time-series prediction, as traditional methods and specialized forecasting models often deliver superior performance on numerical tasks. The group agreed that GenAI and Agentic AI should primarily be used for interpreting demand drivers, communicating uncertainty, structuring response options, and coordinating mitigation actions, rather than replacing dedicated forecasting techniques. This distinction supported the maintenance of both a Demand Agent and a Forecast Agent, each with a defined role: the Forecast Agent addresses prediction, uncertainty, and baseline generation using quantitative methods, while the Demand Agent focuses on anomaly interpretation, early warning, and downstream mitigation to reduce volatility.

Through systematic evaluation of use cases, the focus group identified certain application patterns as particularly promising for disruption readiness. External constraint monitoring, including regulations, sanctions, trade restrictions, and geopolitical developments, was consistently regarded as both highly feasible and impactful. Participants noted that monitoring these domains requires large-scale text comprehension, synthesis, and mapping of unstructured information to specific exposures, which aligns

with the capabilities of Large Language Models. The group concluded that agents capable of continuously interpreting such information and linking it to concrete supply chain risks could help prevent sudden loss of access to critical markets or routes, thereby providing direct resilience benefits. This assessment supported prioritizing the Regulation Agent.

Supplier capability mapping was identified as both technically feasible and operationally valuable. Participants discussed the potential to extract information from unstructured documents, such as contracts, capability statements, and technical specifications, to construct detailed maps of actual supplier capabilities. In the event of disruptions, such agents could identify alternative suppliers capable of fulfilling specific orders based on concrete capabilities, rather than relying on generic classifications. This approach is especially beneficial when supplier information is dispersed across multiple documents and teams. The group's assessment supported the Supplier Capability Agent as a key candidate for resilience-oriented design.

Early detection and amplification damping, particularly regarding demand anomalies, received strong support from participants. They emphasized the significant resilience benefits of early detection, including the prevention or mitigation of amplification dynamics associated with the Bullwhip Effect. The group noted that the greatest leverage occurs upstream: early identification of small anomalies and proportionate responses can substantially reduce downstream volatility in factories, warehouses, and retail. Two broader supply chain insights were repeatedly emphasized. First, the Bullwhip Effect was identified as a central dynamic that resilience interventions should target, with upstream actions offering the highest potential impact. Second, participants highlighted limited transparency beyond Tier 1 suppliers as a fundamental constraint on the effectiveness of Agentic AI in this context.

The focus group also identified areas where Agentic AI is not well suited. Maintenance planning was generally viewed as inappropriate for Agentic AI in its core optimization aspect, as participants considered it primarily a mathematical optimization problem best addressed with established techniques. While Agentic AI may assist with complementary tasks, such as analyzing unstructured maintenance logs, the group did not consider autonomous agent reasoning essential for the primary planning function.

Participants also advised caution regarding overly generic concepts. A generic 'insights agent' producing undefined insights was considered too vague. However, the group endorsed a more focused, resilience-oriented interpretation: an agent designed as a bounded diagnosis-and-options function that (i) synthesizes weak signals and disruption evidence into structured problem statements, (ii) explains likely drivers and constraints using traceable sources, and (iii) generates concise sets of plausible mitigation options within defined decision spaces for human evaluation. Under this definition, participants distinguished such an agent from a generic reporting layer, positioning it as a targeted component that transforms fragmented information into actionable explanations and options during disruption response.

Participants consistently emphasized structural barriers that limit the potential of Agentic AI. Lack of supply chain transparency beyond Tier 1 suppliers was identified as a critical constraint. The group noted that if organizations lack visibility into their Tier 2 or Tier 3 suppliers, even advanced Agentic AI systems may be unable to anticipate disruptions such as sub-supplier bankruptcies or localized catastrophes. In these cases, participants concluded that limited visibility fundamentally restricts what Agentic AI can achieve for resilience, regardless of technical sophistication.

Portfolio Approach and Recommended Agent Shortlist

Through collaborative exercises and plenary discussions, the focus group reached consensus on a clear path forward: instead of developing a single comprehensive resilience solution, the recommended approach is to implement a portfolio of bounded agents, each with specific triggers, defined decision spaces, and measurable outputs. Participants consistently emphasized that value creation depends on precise operational scoping. The rationale for prioritizing certain agents was that they either reduce exposure to external shocks by monitoring rapidly changing constraints, dampen amplification dynamics such as demand volatility and the Bullwhip Effect, or enhance recovery options during disruptions by revealing alternative capacity and capabilities beyond Tier 1 suppliers.

The focus group's final shortlist comprised five agents considered most relevant and actionable for resilience: Demand Agent, Forecast Agent, Insights Agent, Regulation Agent, and Supplier Capability Agent. The overarching conclusion was that building disruption readiness into Agentic AI necessitates recognizing that preparation for unprecedented events cannot depend solely on historical learning. Instead, it requires deliberate system design that ensures crisis behavior is testable, actions are bounded, and human oversight remains significant, even in the face of novel disruptions.

7.2. Thematic Analysis of Follow-up Interviews

In addition to the focus group, two follow-up interviews were conducted with participants to gain further insights into preparing Agentic AI to handle unexpected situations and to explore in greater depth the topics that emerged during the focus group and interviews. The full thematic coding for each interview is provided in Appendix D.

Both participants converged on several key principles for preparing Agentic AI to handle unexpected situations. Enterprise AI implementation was characterized as a configuration challenge rather than foundation-model training, involving mapping business processes to specialized agent networks with explicit tool permissions and orchestrated workflows. Both participants emphasized that predefined workflows should take precedence over free-form autonomy, with bounded agents operating within clearly defined decision spaces. Reliable outputs require multiple safeguards: Retrieval-Augmented Generation and tool connectivity for contextual grounding, alongside guardrails including hard constraints, outcome validators, and neuro-symbolic control. Practical failure modes include over-complex outputs on unfamiliar tasks, hallucination of private information, and performance degradation near context window limits.

For disruption handling, both participants stressed differentiating genuine environmental shifts from model errors. Synthetic scenario generation can prepare systems for plausible failure modes, though practitioners must use domain expert validation to ensure scenarios capture real-world complexity and avoid overfitting. Autonomy boundaries must align with task criticality: non-critical tasks can tolerate fail-and-reset cycles, while critical tasks require stop-on-doubt escalation and redundant agent consensus mechanisms. High-impact contexts demand out-of-distribution recognition and safe defaults that prioritize refusal over speculation. All unexpected cases should be systematically logged for continuous improvement.

Both participants emphasized continuous validation through KPI-based performance comparisons and parallel execution in shadow modes, defining robustness as rapid adaptation capacity following disruptions rather than initial perfect performance. Pragmatic priorities include two-step detection-response

architectures, preferring deterministic methods when feasible, and minimizing complexity. These interviews reinforced the focus group findings: enterprise Agentic AI should be configured rather than retrained, narrowly scoped, equipped with layered safety mechanisms calibrated to task criticality, and continuously validated. The consistent emphasis on conservative fallback strategies and systematic logging supports the view that disruption readiness is fundamentally a system design challenge.

7.3. Concluding Remarks

Preparation extends beyond programming AI agents for predefined tasks and includes equipping them to handle unforeseen situations. Resilience in Agentic AI relies on three primary foundations: synthetic data that generates realistic disruption scenarios, digital twins that offer safe environments for learning, and epistemic-aware methods that allow agents to recognize the limits of their knowledge. The Airbus case illustrates the practical effectiveness of these principles. The aviation industry employs rigorous testing regimes for automation due to the significant consequences of failure. These regimes assess both system capabilities and responses to adverse scenarios. Follow-up interviews highlighted the importance of detection mechanisms for out-of-distribution events, clear handling protocols, and systematic logging of unexpected cases as essential components for robust agent deployment. These observations yield several practical insights. Initial deployment should prioritize simple, reliable tools before introducing complex multi-agent systems. Detection processes should be distinct from response mechanisms: deterministic methods are suitable for structured data, whereas generative AI is more appropriate for unstructured or ambiguous situations. Establishing feedback loops ensures that each unexpected event becomes a learning opportunity for future iterations. In this context, preparation for rare events constitutes a core resilience capability. It is not a one-time setup but an ongoing process of equipping agents for the unknown, testing them in controlled environments, and continuously enhancing their ability to recognize and respond to novel situations.

8

Framework for Resilience Evaluation and AI Method Choice, and Eight Agentic AI Governance Pillars

This chapter translates the research findings into a practical, repeatable framework for two critical decisions: (i) whether a proposed concept constitutes a credible resilience intervention, and (ii) which level of automation is appropriate and consequently whether to deploy GenAI or Agentic AI. While earlier chapters established what these technologies are, how they fail, and which organizational and governance constraints govern their safe adoption, this chapter converts those interview insights into explicit decision logic for determining where to deploy which technology across supply chain decision points. The framework operates in two deliberate dimensions. The first dimension applies resilience evaluation guidelines, synthesized from supply chain expert interviews and literature, to assess whether a concept credibly enhances disruption detection, mitigation, and recovery at a specific decision point. The second dimension introduces Automation choice, derived from AI expert interviews and literature, to determine the appropriate level of capability and autonomy for that decision point. By applying these dimensions sequentially, the framework ensures that candidate concepts are both technically feasible and resilience-worthy: it first evaluates resilience contribution, then selects the suitable AI method and autonomy profile.

8.1. Framework and Dimensions Structure

The framework utilizes two dimensions to assess deployment feasibility and determine the appropriate automation configuration. The resilience dimensions (R1–R5) assess whether a proposed concept provides defensible resilience improvements at a specific SCOR decision point. Subsequently, the choice dimensions (D1–D4) calibrate the suitable level of automation and autonomy for that decision point. This sequential approach ensures that candidate concepts are both resilience-worthy and technically governable prior to implementation.

The framework is structured around decision points, which are defined as specific locations within the Supply Chain Operations Reference (SCOR) model where critical choices directly influence supply chain resilience. A decision point represents a discrete junction at which information is processed, alternative courses of action are assessed, and actions are selected or implemented. The framework is designed for application at each significant decision point within a supply chain. Given that decision points vary in complexity, data requirements, time sensitivity, and risk exposure, the framework is tailored to each decision point. It may recommend different artificial intelligence (AI) methods and levels of autonomy depending on the specific decision under consideration. The optimal deployment of AI and the appropriate degree of automation can vary across SCOR decision points. Therefore, the framework does not prescribe a uniform AI solution for the entire supply chain. Rather, it assists practitioners in identifying the most suitable AI application for each decision point, resulting in a customized automation landscape that aligns with the distinct characteristics and resilience requirements of each decision.

Prior to applying the framework's evaluation dimensions, practitioners must explicitly identify and define the decision point under consideration. This process should specify: (i) the SCOR process category (Plan, Source, Make, Deliver, Return, Enable), (ii) the specific decision or action to be automated, (iii) the decision's scope, frequency, and current execution method, (iv) the stakeholders responsible for decision outcomes, and (v) the information inputs and action outputs involved. This preliminary scoping ensures that subsequent resilience evaluation and AI method selection occur at the appropriate level of granularity and with clearly defined boundaries for automation. The framework is implemented in a sequential manner:

- **Dimension 0 – Decision point identification.** The SCOR decision point that necessitates resilience intervention is clearly identified and defined, specifying its process category, current execution method, stakeholder ownership, and information flows. These specifications delineate clear boundaries for evaluation.
- **Dimension 1 – Resilience fit (R1–R5).** The Concept denotes a particular decision-making approach, which may or may not incorporate artificial intelligence. Its effectiveness is evaluated according to its capacity to enhance resilience. *"Is this a good resilience intervention?"*
- **Dimension 2 – AI method choice (D1–D4).** Only concepts that demonstrate a credible contribution to resilience advance to the method selection stage. The decision point is evaluated based on orchestration requirements, the nature of the work, the alignment between risk and autonomy, and both organizational and technical readiness. *"Given this resilience concept, to what extent is automation feasible?"*

The two dimensions are distinguished by a single criterion. Dimension 1 assesses the concept independently of the artificial intelligence method used for implementation. Dimension 2 evaluates whether the decision point is prepared for a specific level of automation. If a concept does not meet the requirements of Dimension 1, it cannot be considered a credible resilience intervention, regardless of the underlying technology. Concepts that satisfy Dimension 1 but not Dimension 2 may still offer value; however, their implementation strategies must be adapted to the conditions identified in Dimension 2. This separation ensures that technological enthusiasm does not override resilience evaluation. A concept must first demonstrate its merit before the framework addresses implementation.

Dimension 1 - Resilience fit

Table 8.1: Resilience fit

Element	Content
R1 – Vulnerability exposure profile	
Main question	*Does the concept explicitly address vulnerability exposures at the identified decision point?*
Literature support	Pettit et al. (2010) define supply chain vulnerabilities as the fundamental factors that render an enterprise susceptible to disruptions, consolidating these into seven measurable dimensions: turbulence, deliberate threats, external pressures, resource limits, sensitivity, connectivity, and supplier or customer disruptions. At the decision-point level, R1 requires an explicit identification of the dominant vulnerability dimensions within the specific context, along with justification. It further assesses whether the concept addresses these particular exposure mechanisms. Concepts are evaluated more favorably when they align with the local vulnerability profile.
Interview Support	The expert interviews offer robust empirical support for this guideline. Turbulence is observed as demand shocks resulting from unexpected promotion changes. External pressures are identified in infrastructure failures, such as canal blockages. Resource limits are evident in distribution center congestion. Deliberate threats encompass theft and security breaches. Supplier disruptions are attributed to unreliable transport that necessitates urgent modal shifts. Connectivity is demonstrated by the propagation of the bullwhip effect from upstream shocks. Sensitivity is also acknowledged, as practitioners report that low inventory levels can escalate minor shortages into significant production stoppages. These results underscore the importance of systematic vulnerability profiling at each decision point.
R2 – Capability portfolio fit and balance	
Main question	*Does the concept deliver a concrete portfolio of resilience capabilities that is balanced against the vulnerability profile, and does the resulting performance avoid both excessive risk (from under-capability) and eroded profitability (from over-capability)?*
Literature support	Pettit et al. (2010) define capabilities as management-controlled attributes that enable an enterprise to anticipate and overcome disruptions and show that resilience improves when capabilities increase and vulnerabilities decrease, but that performance is maximized when both are *balanced* (the ”Zone of Resilience”): excessive vulnerabilities relative to capabilities yield excessive risk, while excessive capabilities relative to vulnerabilities erode profits. Accordingly, R2 evaluates whether the concept strengthens a defensible subset of capability dimensions and matches them to the decision point’s dominant vulnerabilities. The capability dimensions include: *flexibility in sourcing*, *flexibility in order fulfillment*, *capacity*, *efficiency*, *visibility*, *adaptability*, *anticipation*, *recovery*, *dispersion*, *collaboration*, *organization*, *market position*, *security*, and *financial strength*. A concept scores higher when it specifies which capability levers it activates, how those levers offset the identified vulnerabilities at the decision point, and why the implied investment level is proportionate to the exposure pattern.

Continued on next page.

Table 8.1 continued.

Element	Content
Interview Support	Expert interviews indicate that practitioners employ multiple capability levers and emphasize the importance of proportionate investment. Structural resilience is established through multi-plant networks, which provide dispersion and flexibility in sourcing, thereby enabling rapid production shifts. Operational resilience depends on promotion visibility and volume reallocation across markets, ensuring visibility and flexibility in order fulfillment. Practitioners utilize explicit threshold rules and metrics to balance efficiency with service risk, and implement mitigation strategies such as production reallocation, alternative sourcing, and safety stock borrowing to enhance recovery and capacity. AI experts further support the balance principle, observing that Agentic AI introduces higher costs due to orchestration and monitoring. They assert that value must be justified by proportionate gains, confirming that capability investments should align with exposure rather than pursue maximal levels.
R3 – Data and IT realism, and visibility impact	
Main question	*Is the concept capable of improving resilience given the actual state of available data and system integration, and does it enhance supply chain visibility - the ability to access and interpret timely, accurate information across the end-to-end supply chain - at the identified decision point?*
Interview Support	Expert interviews consistently indicate that data quality and system integration are critical constraints for Agentic AI deployment. Practitioners report that upstream risk tools often encounter incomplete, inconsistent, or poorly integrated data, which necessitates reliance on probabilistic estimates instead of comprehensive traceability. Skepticism regarding Agentic AI is primarily attributed to insufficient data and frequent disconnections among IT systems. Experts emphasize that resilience depends on robust IT infrastructure and continuous monitoring. AI specialists identify data maturity, including a thorough understanding of data sources, definitions, and quality, as a primary condition for readiness. They further note that poor data inputs inevitably result in poor outputs, regardless of Agentic AI validation. Consequently, this guideline assesses the feasibility of the concept under actual data conditions and aims to achieve meaningful improvements in visibility, rather than presuming ideal data availability.
R4 – Governance, incentives, and collaboration fit	
Main question	*Is the concept aligned with the real governance and incentives, and does it support collaboration instead of fighting it?*
Interview Support	Expert interviews indicate that resilience interventions are influenced by incentive structures, collaboration constraints, and automation governance requirements. Practitioners note that internal decision-making is undermined when teams prioritize individual metrics, resulting in friction and reduced profitability. External collaboration improves resilience when partners engage in forecasting and provide order visibility; however, organizations are often unwilling to share supplier data, viewing it as a source of competitive advantage. Regarding automation governance, practitioners anticipate that Agentic AI will execute reallocations within established parameters, such as financial thresholds and approval workflows. AI experts assert that autonomy should be limited by governance mechanisms, including policies, defined autonomy levels, access controls, and accountability structures. This guideline assesses alignment with existing governance frameworks and examines whether the proposed autonomy preserves accountable decision rights.
R5 – Cross-functional integration at the decision point	
Main question	*Does the concept ensure that the identified decision point is integrated into the broader supply chain process - enabling coordinated, cross-functional behaviour rather than isolated actions that generate disconnected alerts and increase planning noise?*

Continued on next page.

Table 8.1 continued.

Element	Content
Interview Support	Expert interviews suggest that AI agents perform most effectively when functioning as part of an integrated system rather than in isolation. Disruptions at any decision point can create cascading effects throughout the supply chain. For instance, a shipping delay at one node may alter inventory levels, which in turn affects production and order fulfillment in other areas. Due to these interdependencies, the concept should clearly demonstrate how the Agentic AI system at the specified decision point coordinates with adjacent functions and processes. This guideline assesses three main criteria. First, it examines whether the concept specifies how agents at the decision point exchange information with upstream and downstream systems using standardized data formats. Second, it considers whether a defined process exists for decision flows triggered at the decision point to propagate appropriately, such as a forecast change prompting adjustments in inventory, production, and shipping. Third, it evaluates whether the system consolidates alerts originating from or arriving at the decision point into prioritized actions for relevant stakeholders, rather than generating excessive notifications. A concept is rated poorly if it fails to account for cross-functional dependencies at the decision point, generates alerts that are not actionable by adjacent systems, or lacks a clear explanation of how coordination with the broader supply chain is achieved.

Mapping R-Guidelines to the Disruption Profile

The five resilience guidelines (R1–R5) demonstrate differing levels of relevance throughout the various phases of a disruption, with each guideline attaining peak importance at distinct points along the disruption profile (Figure 2.2). Together, these guidelines provide comprehensive coverage from preparation (Phase 1) through to long-term impact (Phase 8), thereby addressing all phases of the disruption.

R1 (Vulnerability Exposure Profile) is most relevant during Phase 1 (Preparation), as it facilitates the identification of system exposures and assesses whether the concept addresses the dominant vulnerability dimensions at the decision point. R1 is also pertinent to Phase 8, where the cyclical learning process informs updates to the vulnerability profile for subsequent cycles.

R2 (Capability Portfolio Fit and Balance) is most applicable during Phase 1 (Preparation), where it assesses whether the concept provides a portfolio of capabilities proportionately balanced against the vulnerability profile, thus avoiding both under-capability and over-capability. R2 is also relevant to Phase 8, as post-disruption evaluation determines whether the capability balance was adequate and informs recalibration of investments for future cycles.

R3 (Data and IT Realism and Visibility Impact) functions as a continuous enabler, with peak importance during Phases 1 to 4. Enhanced visibility across the supply chain supports early detection during preparation and the initial response, and enables real-time assessment of the magnitude and propagation of the disruption during Phases 4 and 5.

R4 (Governance, Incentives, and Collaboration Fit) is structurally established in Phase 1 but becomes critical during Phases 3 to 7, when misaligned incentives and collaboration barriers can directly undermine coordinated response and recovery.

R5 (Cross-Functional Integration) aligns most closely with Phases 3 to 7, during which cascading effects propagate through interconnected functions and cross-functional coordination determines response effectiveness. R5 is also relevant to Phase 8, as the long-term learning loop depends on consolidated

cross-functional feedback.

Structurally, R1 and R2 follow a design-then-activate pattern, shaping preparation and determining available resources when disruption occurs. R3 and R4 serve as enabling conditions that must be established prior to disruption and are actively tested throughout the response. R5 is most operationally visible during response and recovery, but must be architecturally embedded during design. Table 8.2 summarizes the primary and secondary phase relevance for each guideline.

Table 8.2: R-guideline mapping to disruption profile phases

Guideline	Primary Phases	Secondary Phases
R1 Vulnerability exposure	Phase 1 (Preparation)	Phase 8 (Long-term impact)
R2 Capability balance	Phase 1 (Preparation)	Phase 8 (Long-term impact)
R3 Data and visibility	Phases 1–4 (Preparation through Initial impact)	Phases 4–5 (Initial impact, Full impact)
R4 Governance and collaboration	Phases 3–7 (First response through Recovery)	Phase 1 (Preparation)
R5 Cross-functional integration	Phases 3–7 (First response through Recovery)	Phase 8 (Long-term impact)

Dimension 2 - Design Prescription

A concept progresses to Dimension 2 only when its aggregate R-score demonstrates a credible contribution to supply chain resilience. At this stage, Dimension 2 does not reconsider the concept's value, as this determination has already been made. Instead, Dimension 2 evaluates the appropriate and feasible level of automation for the specific decision point. The relevant inputs also shift: Dimension 1 relies on the vulnerability profile, capability dimensions, and supply chain structure, whereas Dimension 2 considers the complexity of task orchestration, the nature of the work, the risk-autonomy ceiling, and organizational readiness.

Table 8.3: Design Prescriptions

Element	Content
D1 – Orchestration versus single interaction	
Main question	Does this resilience use case require orchestration across *multiple tools, systems, or process steps* (plan–act–verify), or is it essentially a single interaction?

Continued on next page.

Table 8.3 continued.

Element	Content
Interview Support	Expert interviews delineate a distinct capability boundary between GenAI and Agentic AI, grounded in orchestration requirements. GenAI is defined as an interaction-level capability appropriate for bounded, single-step tasks. In contrast, Agentic AI coordinates multiple steps, tools, and resources through a plan–act–verify loop. The principal distinction is agency: Agentic AI functions through a closed-loop process involving goal-setting, action execution using external tools, and outcome evaluation, rather than producing a single, one-time response. This boundary is further supported by observed supply chain application patterns. Experts note that effective supply chain performance relies on orchestrating fragmented data and workflows, such as reconciling data across systems and continuously monitoring external signals. These capabilities align with the architectural strengths of Agentic AI. Agents provide value when they manage defined duties end-to-end, rather than generating isolated responses. D1 therefore evaluates whether orchestration across multiple tools, systems, or process steps constitutes a structural requirement of the use case, which would necessitate Agentic AI design. If orchestration is discretionary, GenAI or deterministic workflows are preferable due to their lower complexity, cost, and risk of failure.
D2 – Nature of work: judgment versus low-creativity workflow	
Main question	Is the core of the task judgment and interpretation, or low-creativity, repetitive data work with clear rules?
Interview Support	Expert interviews consistently differentiate between two categories of work that necessitate distinct AI methodologies. Judgment-dominant tasks rely on contextual interpretation, entail significant subjectivity, and require human accountability for final decisions. In contrast, execution-dominant routine tasks generate value by systematically applying structured procedures to large datasets governed by explicit rules. GenAI is most suitable for tasks that demand substantial judgment or involve high-stakes outcomes, where human oversight remains critical for final decision-making. Agentic AI is more appropriate for repetitive, low-creativity tasks that process extensive volumes of unstructured data and require automated analysis to directly initiate actions. Agentic AI systems exhibit superior performance in executing well-defined, repetitive tasks at scale without fatigue, whereas high-stakes exceptions should be escalated for human intervention rather than automated. D2 therefore assesses whether the task's core value lies in supporting human reasoning without delegating execution (favouring GenAI or advisory configurations), or in bounded, repeatable workflow execution at scale (where Agentic AI's orchestration overhead can be economically defended).
D3 – Risk level and autonomy fit	
Main question	Given the impact of a bad decision here, how much autonomy can we tolerate?

Continued on next page.

Table 8.3 continued.

Element	Content
Interview Support	Expert interviews identify model-level errors, such as hallucination and poor mathematical computation, as well as data vulnerabilities, including polluted training data, prompt injection, and garbage-in/garbage-out, and Agentic AI coordination risks, such as agent contradiction, compromised components, and handoff failures, as typical operating conditions rather than rare anomalies. Agentic AI architectures do not eliminate these errors; even with validated intermediate steps, workflows may still converge on incorrect results. In Agentic AI workflows, risks are further amplified by extreme agency, where a single error can propagate into multiple subsequent failures. Accordingly, the degree of autonomy should be aligned with the severity of potential consequences and the system's capacity for containment. Oversight requirements should be based on the risk level of the task and relevant contextual factors, rather than solely on system classification. For tasks with severe potential consequences, robust human oversight is essential, whereas lower-risk tasks with proven performance may warrant greater autonomy. This calibration is implemented through defined autonomy levels, approval rights, hard stops, escalation protocols and rollback mechanisms. D3 represents the maximum permissible level of autonomy at a decision point, taking into account the potential impact of errors, the amplification risks inherent in Agentic AI architectures, and the associated oversight requirements.
D4 – Organizational and technical readiness	
Main question	Does this process have the data, integration, ownership, and safeguards needed to support something beyond GenAI?
Interview Support	Expert interviews indicate that the primary constraint on Agentic AI deployment is not model capability, but rather the organization's ability to specify, monitor, and control the system in practice. Two foundational readiness conditions are identified: data maturity, which involves understanding data sources, definitions, and quality while avoiding unreliable inputs, and employee AI maturity, defined as the skills necessary to use AI effectively beyond basic tasks. Three essential infrastructure prerequisites are identified: high-quality, reliable data; robust system interfaces that prevent fragile integrations; and dedicated testing environments for safe verification prior to deployment. Safety should be embedded from the outset through data protection, exception prevention, and mechanisms for post-incident learning. Operational safeguards include explicit accountability for outcome ownership, comprehensive monitoring and logging, deterministic checks, and predefined fallback options with rollback mechanisms. Initial deployments should focus on low-risk, well-defined tasks with established reference processes, validating both performance and safeguards before incrementally increasing autonomy. D4 thus assesses whether the environment, considering data quality, system integration, process ownership, and safety architecture, can support applications beyond advisory GenAI, regardless of whether an agent could theoretically be constructed.

8.2. Theoretical AI Foundation and Method Mapping

Parasuraman et al. (2000) established that automation is not binary but exists along a continuum requiring deliberate calibration. Their foundational argument is that "automation does not merely supplant but changes human activity and can impose new coordination demands on the human operator." Critically, they argue that "the human performance consequences of particular types and levels of automation constitute primary evaluative criteria for automation design"(Parasuraman et al., 2000), meaning the appropriate level should be determined by human-system performance outcomes rather than technical capability alone. The ten-level scale characterizes the progressive transfer of control from human to

computer. Table 8.4 summarizes these levels and their implications for human oversight.

Table 8.4: Levels of Automation scale retrieved from (Parasuraman et al., 2000)

LoA	Description	Human–Automation Interaction
1	No assistance	The human does everything manually; the automation offers no assistance.
2	Full alternatives	The automation offers the complete set of decision alternatives; the human decides.
3	Narrowed options	The automation narrows options to a few possibilities for the human to choose from.
4	Single suggestion	The automation suggests one alternative; the human retains authority to accept or reject.
5	Execute if approved	The automation suggests one alternative and executes it if the human approves.
6	Veto window	The automation executes automatically unless the human vetoes within a limited time.
7	Inform the human after	The automation executes and informs the human only after the fact.
8	Inform the human if decides to	The automation executes and informs the human only if it decides to.
9	Inform the human if asked	The automation decides everything and informs the human only if asked.
10	Full autonomy	The automation decides and acts autonomously, ignoring the human entirely.

Mapping AI Architectures to Automation Levels

The framework assigns GenAI and Agentic AI to specific automation levels according to their unique architectural capabilities. This mapping, supported by interview and focus group evidence from previous chapters, demonstrates how various AI architectures correspond to increasing levels of automation within the information processing and decision-action continuum.

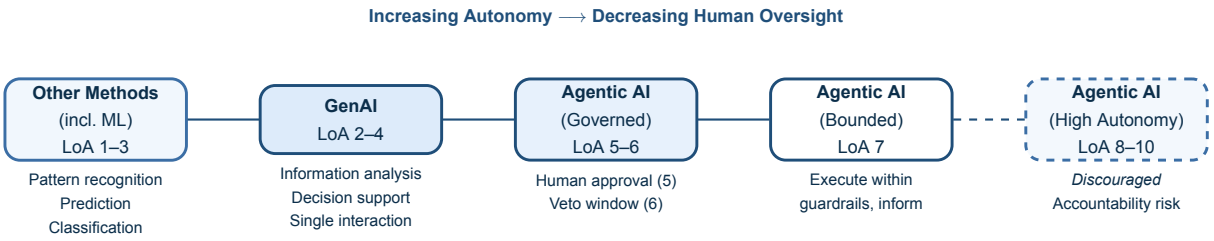


Figure 8.1: Mapping of AI architectures to Levels of Automation (LoA), showing the progression from traditional methods through generative to Agentic systems along the autonomy continuum.

Other Methods, including traditional Machine Learning (ML), occupy the lower end of the automation continuum (LoA 1 to 3). Approaches such as supervised classification, regression, and cluster-

ing automate pattern recognition and prediction tasks. Human operators are responsible for interpreting outputs and determining subsequent steps. **GenAI** systems operate at the information analysis stage (LoA 2–4), where they synthesize data and recommend actions through single-interaction outputs. These systems enhance human cognition in an advisory role, without executing decisions or coordinating workflows, thereby preserving human decision-making authority. **Agentic AI** extends beyond GenAI by decomposing goals, invoking external systems, and autonomously executing multi-step workflows. This capability situates Agentic AI at LoA 5 and higher, with specific configurations selected based on organizational risk tolerance. **Governed Agentic AI** ensures robust human oversight through explicit approval at LoA 5 or time-limited veto periods at LoA 6. In these configurations, agents prepare executable action packages, while humans retain final authorization authority. **Bounded Agentic AI** introduces human-on-the-loop oversight, permitting agents to execute actions autonomously within predefined constraints and to notify supervisors after execution. This approach necessitates increased risk tolerance, robust rollback mechanisms, and validated action boundaries, marking a substantial governance shift from LoA 5–6. **High-Autonomy Agentic AI** permits decisions and actions with minimal human intervention.

8.3. Dimensions Interaction Matrix

The Dimensions Interaction Matrix (Figure 8.2) operationalizes the two-dimension evaluation framework by mapping R-Dimension scores, which reflect resilience contribution, against D-Dimension scores, which represent implementation feasibility and autonomy fit. This matrix divides the decision space into six distinct zones, each prescribing specific actions according to resilience value and organizational readiness.

Zone 1: Reject – Neither Valuable Nor Governable. Concepts located in the bottom-left quadrant fail to meet both criteria and should be rejected. Low R-Dimension scores demonstrate that the concept does not adequately address vulnerability profiles or yield measurable resilience outcomes. Low D-Dimension scores reflect insufficient organizational readiness or a misalignment between task characteristics and automation capabilities.

Zone 2: Feasible but Not Valuable. Concepts in this zone achieve high D-Dimension scores but low R-Dimension scores, making them technically feasible but lacking resilience value, reflecting a technology-driven rather than problem-driven approach. While implementation may appear advantageous, deploying excessive capabilities without addressing validated vulnerabilities diminishes profitability and does not enhance resilience (Pettit et al., 2010). Implementation for resilience purposes is not recommended unless there are demonstrable productivity gains beyond the resilience mandate.

Zone 3: Valuable but Not Governable. This zone is characterized by high R-Dimension scores but low D-Dimension scores, indicating strong resilience value but insufficient organizational readiness for autonomous execution. The recommended configuration is Levels of Autonomy (LoA) 1 to 4, in which GenAI provides advisory support while human operators retain decision authority. This configuration is appropriate when data integration is incomplete, error profiles carry significant consequences, or tasks require contextual judgment. GenAI synthesizes information, generates scenario analyzes, and recommends actions without executing them, thereby capturing resilience value while mitigating governance risks and maintaining human oversight.

Zone 4: Agentic AI – Governed Autonomy. Concepts achieving high R-Dimension scores and moderate D-Dimension scores are suitable for Agentic AI deployment at LoA 5 and 6. At LoA 5 (Execute if Approved), agents prepare executable action packages that require explicit human approval. At LoA

6 (Veto Window), agents act automatically unless a veto is issued within a specified timeframe. This approach is appropriate for time-sensitive resilience responses where delays result in significant costs, yet oversight remains essential.

Zone 5: Agentic AI – Bounded Autonomy. High scores on both the R-Dimension and D-Dimension warrant increased autonomy at LoA 7. Concepts in this zone exhibit strong resilience contribution, advanced data integration, appropriate risk-autonomy alignment with comprehensive guardrails, and execution-focused workflows. Agents operate autonomously within predefined boundaries and notify stakeholders after execution. Guardrails, such as financial thresholds and exception triggers, maintain safety within bounded autonomy. All actions are logged to ensure traceability and enable intervention if necessary.

Zone 6: High Autonomy (Discouraged). The framework establishes an upper limit on permissible autonomy at Level of Autonomy (LoA) 7, regardless of R-Dimension and D-Dimension scores. This ceiling is based on a distinct evaluative principle: the preservation of human accountability, situational awareness, and intervention capability. Evidence from interviews and focus groups consistently demonstrates that LoA 8 to 10 undermine these properties in ways that cannot be offset by high resilience or strong organizational readiness. This boundary represents a categorical, rather than incremental, discontinuity, a distinction that has a structural parallel in the SAE J3016 driving automation taxonomy (SAE International, 2021). Within this taxonomy, the transition from Level 3 (Conditional Automation) to Level 4 (High Automation) signifies the point at which human fallback responsibility is structurally removed. At Level 3, a fallback - ready user must remain present and attentive to resume vehicle control when the system is unable to do so. At Level 4 and above, the Automated Driving System assumes full responsibility for both the dynamic driving task and fallback execution, and no human intervention is required or expected (SAE International, 2021). The present framework applies this same reasoning: crossing from Zone 5 into Zone 6 is not a small step but a fundamental shift, where human accountability does not simply decrease - it disappears entirely. Consequently, the governance ceiling serves as a strict constraint on the matrix output. Even when both dimensions scores would allow for greater autonomy, the framework limits implementation to LoA 7 with robust safeguards.

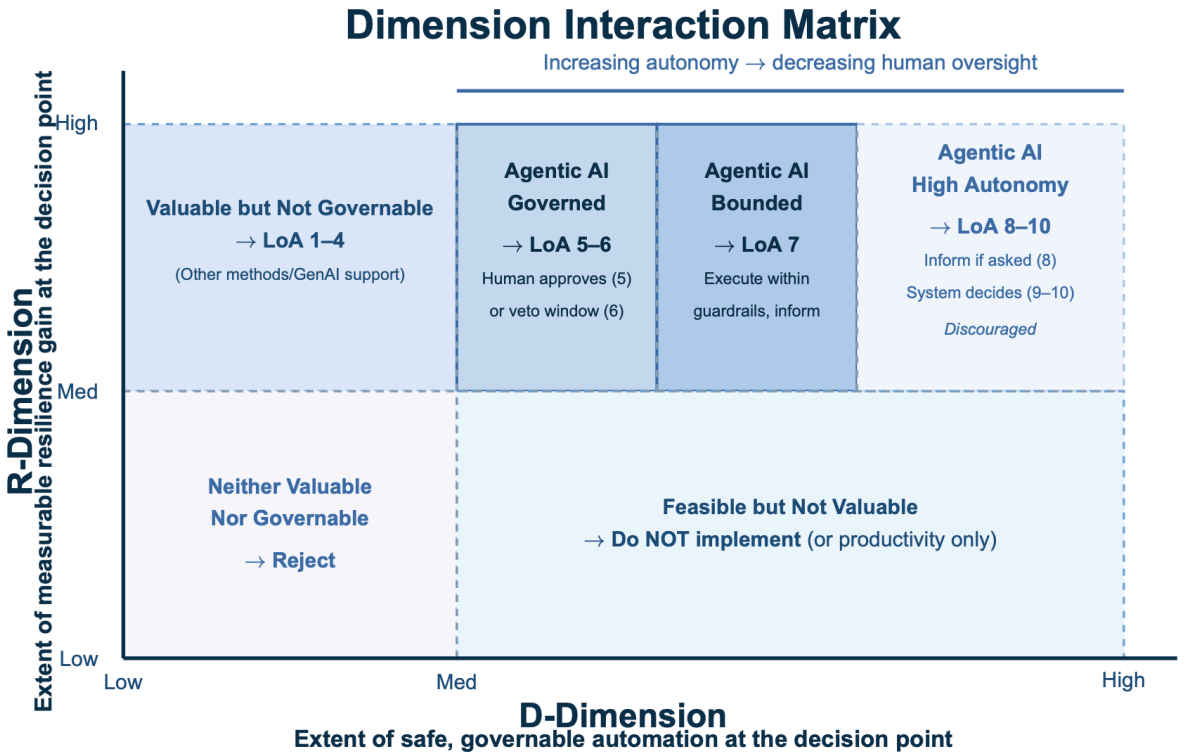


Figure 8.2: Dimension Interaction Matrix mapping resilience impact against automation governability to classify concepts into six decision zones and prescribe corresponding LoA-aligned actions

8.4. Training Capabilities Required by Levels of Automation

Section 8.1 describes a framework that uses two dimensions to assess deployment feasibility and determine the appropriate automation configuration. For each Level of Automation (LoA) range, specific training capabilities are necessary because the system's outputs transition from advisory support (LoA 1–4), to execution intent under human control (LoA 5–6), and finally to autonomous execution prior to review (LoA 7). Higher LoA configurations therefore require increasingly robust training strategies, including expanding scenario coverage beyond historical operations, validating multi-step behavior in controlled environments when sequential decisions are critical, and implementing uncertainty-aware safeguards to prevent overconfident actions in unfamiliar situations. This section details which training capabilities are required, recommended, or discretionary for each zone and its associated LoA range, and validates this mapping against the automation assurance practices of Airbus, where preparedness for rare events is considered a requirement rather than an aspiration.

Zone-Specific Training Requirements

Chapter 7 identified three interdependent training capabilities for Agentic AI architectures: *synthetic data generation*, which broadens scenario coverage beyond historical operations; *digital twin environments*, which facilitate safe learning and validation of sequential decision policies; and *epistemic-aware methods*, which support graceful degradation when the system encounters conditions outside its learned envelope. The necessity of these techniques varies across the zones of the Interaction Matrix. Their applicability is determined by the autonomy level assigned to each zone and the potential severity of consequences if the system operates confidently in unfamiliar conditions. Table 8.5 aligns

each training capability with the corresponding matrix zones and provides the operational rationale for its inclusion.

Table 8.5: Training capability requirements by Interaction Matrix zone

Zone	LoA	Synthetic Data	Digital Twin	Epistemic-Aware Methods
Zone 1 Reject	n/a	Not applicable	Not applicable	Not applicable
Zone 2 Feasible, not valuable	n/a	Not applicable	Not applicable	Not applicable
Zone 3 Valuable, not governable	1–4	**Low priority.** The use of synthetic records enables coverage of exception cases and disruption narratives, thereby enhancing the quality of advisory outputs.	**Low priority.** At this level of autonomy, the human operator maintains complete decision-making authority, and sequential learning is not required.	**Low priority.** Flagging low-confidence recommendations is beneficial, however, the human decision-maker remains the primary safeguard against actions based on uncertain information.
Zone 4 Governed autonomy	5–6	**High priority.** The agent generates executable action packages (LoA 5) or initiates actions unless a veto is applied (LoA 6). Scenario coverage should encompass not only historical operations but also novel events to ensure that proposed actions are rigorously evaluated against situations absent from historical data.	**Medium priority.** Digital twins enable the validation of complex action sequences before deployment and provide a controlled environment to evaluate the performance of proposed actions under cumulative disturbances.	**High priority.** At LoA 5 to 6, the agent's recommendations are intended for execution. Implementing out-of-distribution detection and uncertainty-triggered escalation, such as switching from LoA 6 to LoA 4 when epistemic uncertainty surpasses a defined threshold, is essential to prevent unreliable yet confident action proposals from being presented to human approvers without sufficient warnings.

Continued on next page.

Table 8.5 continued.

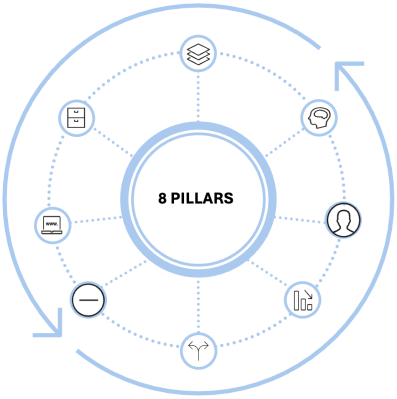
Zone	LoA	Synthetic Data	Digital Twin	Epistemic-Aware Methods
Zone 5 Bounded autonomy	7	**High priority.** The agent operates autonomously within predefined guardrails and provides post hoc notifications. Any gap in scenario coverage constitutes a regime in which the agent may act with unwarranted confidence. Comprehensive synthetic scenario libraries, encompassing compound and cascading disruptions, are necessary to ensure that guardrail boundaries are robust.	**High priority.** Sequential decision policies should be developed and validated in controlled environments prior to autonomous deployment. Implementing domain randomization within the digital twin is essential to mitigate the sim-to-real transfer gap.	**High priority.** For autonomous execution without prior human approval, the system must independently detect when it operates outside of learned conditions. Mechanisms such as abstention gates, selective prediction, and dynamic autonomy scaling, which reduces the level of autonomy from LoA 7 to LoA 5 or lower under high uncertainty, serve as critical safeguards to prevent error propagation during multi-step autonomous actions.
Zone 6 High autonomy (discouraged)	8–10	The framework restricts permissible autonomy to Level of Autonomy (LoA) 7, irrespective of the extent of training investment. Neither comprehensive scenario coverage, high-fidelity digital twins, nor advanced epistemic calibration can offset the reduction in human accountability, situational awareness, and intervention capability that defines LoA 8 to 10. Enhanced training capabilities cannot supersede this governance limitation.		

This mapping demonstrates that training capabilities are not generic investments but rather specific prerequisites whose necessity increases with the prescribed autonomy level. At LoA 1 to 4 (Zone 3), the human decision-maker assumes responsibility for system uncertainty, and training capabilities primarily enhance information quality. At LoA 5 to 6 (Zone 4), the agent's outputs possess execution intent, necessitating scenario-validated proposals and mechanisms for uncertainty-aware escalation. At LoA 7 (Zone 5), where execution occurs prior to human review, all three capabilities become essential, as any gap in the system's learned envelope may result in uncontained autonomous error. This progression clarifies the rationale for the absolute governance ceiling at LoA 7: beyond this point, the system's ability to recognize its own limitations, regardless of engineering quality, cannot replace the structural assurance provided by human intervention, which is absent at LoA 8 to 10. Practically, LoA 5 to 6 configurations require synthetic scenario generation and epistemic-aware escalation mechanisms, as outputs possess execution intent and must be rigorously tested and qualified for safety under novel conditions. LoA 7 configurations further necessitate digital twin environments for validation, since autonomous execution prior to review requires evidence that sequential behavior remains safe amid diverse and compounding disruptions. Consequently, training capabilities serve as the operational link between autonomy selection and the safe, governed deployment of autonomous systems.

8.5. Transferrable Principles - Eight Pillars

This section synthesizes findings from previous research phases into eight transferable design patterns for governing Agentic AI in supply chain resilience applications. These pillars result from a comprehensive integration of all prior research activities. Expert interviews, focus groups, the case study, and the Airbus cross-industry analogy each contributed distinct insights. These contributions encompass the identification of governance gaps and failure modes, the definition of oversight boundaries, and the provision of established architectural mechanisms from a mature, safety-critical automation domain.

Each pillar represents a convergence of a governance need, identified through interviews or focus groups, with a validated mechanism, which is then translated into a concrete design pattern for Agentic AI. Figure 8.3 provides an overview of these eight pillars.



Layered control stack (planning -> guidance -> execution)	Separate “What to do?” (planning) from “How to do it?” (execution) so local errors don’t cascade system wide
Command / Monitor pattern	One agent commands the workflow; an independent supervisor monitors and can block/stop when outputs conflict or risk rises
Managed vs Selected modes	Use Managed mode for routine operations and switch to selected mode when data integrity is uncertain.
Graceful degradation	When uncertainty or inconsistency increases the system has a graceful degradation: Execute -> Recommend-only -> Manual Control.
Stabilization-First under surprises	If outcomes diverge from expectations, prioritize containment and stability with simple, transparent rules before re-engaging in richer automation.
Digital Supply Chain Twin as “Iron Bird”	Test extreme behaviours safely before real deployment
Digital Black Box	Log tool calls, actions and anomalies so post-event reviews improve requirements and safeguards.
Parallel Architecture and Redundant Decision-Making	Independent parallel systems cross-check each other and use voting to isolate faults and fall back safely.

Figure 8.3: Eight transferable design patterns for governing Agentic AI in supply chain resilience applications

- **Pillar I: Layered Control Stack**

  The need for a layered control architecture emerged during multiple research phases. Experts highlighted the significance of an integrated system that incorporates governance, safeguards, and exception management. The focus group further recommended deploying secondary agents to oversee primary agents. This principle is exemplified in Airbus flight systems, which utilize a three-layer architecture that separates planning, guidance, and execution, with each layer validating inputs before action. Applying this model to Agentic AI results in three analogous tiers: Strategic Planning Agents, responsible for defining high-level objectives; Tactical Guidance Agents, which convert strategy into actionable steps; and Operational Execution Agents, which directly interact with the system. Each layer validates its inputs against established constraints before proceeding. This structure mitigates cascade risks in complex supply networks by detecting planning errors prior to execution and preventing operational issues from affecting strategic decisions.

- **Pillar II: Command/Monitor Pattern**

  The principle of separating execution from independent verification was consistently identified in both interviews and the focus group. Experts supported the concept of ”agency over agency” and recommended the use of distinct evaluator agents for cross-validation. The focus group further formalized the role of supervisory agents tasked with detecting anomalies and verifying

approvals before execution. This separation is analogous to the paired command and monitor channels in flight systems, which use distinct logic and data sources to cross-validate outputs. In the context of Agentic AI, a Command Agent executes decisions, while an independent Monitor Agent, operating with separate data, models, and logic, can halt execution if conflicting outputs are detected. This architecture exposes hidden failures that single-channel systems may miss and provides early warnings when outputs diverge, allowing for investigation before minor errors escalate.

- **Pillar III: Managed vs. Selected Modes**

  The theme of calibrating autonomy according to context was consistently identified during the discussion. Experts agreed that autonomy should be proportional to risk. The focus group established governance boundaries that define two distinct operational modes, while supply chain specialists conceptualized oversight as a spectrum based on criticality rather than a binary choice. A relevant operational example is the Airbus Flight Management and Guidance System (FMGS), which provides a managed mode, where the system follows flight-plan-derived targets such as speed, altitude, and trajectory, and a selected mode, where pilots directly input parameters. Flight crews typically switch to selected mode when predictability or data integrity is compromised, as direct inputs enable rapid adjustments. For Agentic AI, this principle suggests a dual-mode design: (1) an automated mode in which the agent executes pre-computed plans and optimized targets, and (2) a direct-control mode in which humans set parameters, overrides, or constraints. This approach maintains the benefits of automation in stable conditions while enabling targeted human intervention when required.

- **Pillar IV: Graceful Degradation**

  The principle that systems should degrade gracefully rather than fail catastrophically was established through multiple research phases. Experts identified escalation, rollback, and predefined fallback options as essential design objectives. The focus group developed context-dependent escalation rules to route issues to the appropriate teams. In flight systems, this principle is implemented by progressively reducing control levels and providing clear alerts. Agentic artificial intelligence (AI) should adopt a similar approach by transitioning from full autonomy to AI-generated recommendations requiring human approval, and ultimately to manual control supported by decision aids. The level of autonomy should adjust automatically based on data quality, model confidence, and system health, with explicit alerts at each transition. This strategy prevents system collapse during failures, maintains safe and controllable decision-making, and enables recovery to full operational capability as conditions improve.

- **Pillar V: Stabilization-First Under Surprises**

  The principle of implementing containment prior to diagnosis is embedded within operational routines, as outlined by supply chain experts, with containment measures preceding root-cause analysis. Nevertheless, the focus group identified a significant gap: when exceptions arise outside established protocols, teams often resort to improvisation rather than following a structured containment-then-diagnosis process. Aviation recovery protocols specifically address this issue by requiring stabilization before diagnosis. Agentic AI systems should incorporate the same sequence: (1) Immediate Containment, (2) Situation Assessment, with diagnosis occurring only after containment, and (3) Gradual Re-engagement. This approach prevents external shocks from propagating through interconnected systems and enables rapid containment prior to engaging in

complex diagnostics that may increase uncertainty.

- **Pillar VI: Digital Supply Chain Twin as “Iron Bird”**

  Simulation-based preparation emerged as a critical requirement based on independent expert interviews, where both AI and supply chain professionals emphasized the need for forward-looking simulation and dedicated testing environments. The second focus group corroborated this finding, achieving consensus on the significance of synthetic scenario creation, edge-case unit testing, and shadow deployment as complementary readiness strategies. The aviation industry’s ”Iron Bird” concept, which utilizes digital models and hybrid rigs to test systems prior to deployment with synthetic data to address edge cases, serves as a practical model for these requirements. Consequently, Agentic AI systems should develop digital twins of the supply chain prior to deployment, generate synthetic disruption scenarios to allow agents to practice failure modes safely, and implement policies only after demonstrating robustness through comprehensive stress tests. These measures prepare systems for rare catastrophic events not present in historical data and support proactive mitigation design rather than reactive crisis management.

- **Pillar VII: Digital Black Box**

  Comprehensive logging and post-event learning have been identified as essential requirements across all research phases. Experts highlight traceability, state logging, and rollback as critical safeguards. Drawing on the flight-recorder principle foundational to aviation’s culture of continuous improvement, Agentic AI systems should automatically record actions, including parameters, and decision rationales such as confidence levels, input data, and considered alternatives. Logs must also capture human interventions, including overrides and approvals, and document anomaly triggers such as state transitions and alerts. Subsequent post-disruption analysis facilitates systematic improvements by translating identified lessons into updated specifications, thereby reducing the likelihood of repeated failures and establishing a controlled process for continuous improvement, as opposed to ad hoc remediation.

- **Pillar VIII: Parallel Architecture and Redundant Decision-Making**

  Experts have identified the risk that a single compromised component can disrupt the entire workflow, leading to the proposal of cross-agent validation. The focus group determined that implementing multi-layered oversight introduces redundancy, thereby mitigating the impact of imperfections in AI outputs. Airbus applies this principle at several levels: two redundant Flight Management and Guidance Computers operate in parallel and cross-monitor each other; three independent Air Data and Inertial Reference Units are cross-validated using majority voting to identify faults; and electrical and hydraulic power systems are separated to ensure that a single fault does not compromise the entire system. When outputs diverge, voting logic detects inconsistencies, enabling the system to deactivate the malfunctioning unit and transition to a degraded operational mode. Applying this approach to Agentic AI for supply chain resilience involves deploying architecturally distinct models or agent implementations that independently generate decisions for the same critical task. A voting or consensus mechanism then compares these independent outputs to determine the actionable result. If outputs exceed acceptable thresholds, execution is paused and the decision is escalated to human review or directed to a fallback process.

The eight pillars integrate findings from expert interviews, focus groups, case studies, and aviation automation experience into actionable design patterns for Agentic AI aimed at enhancing supply chain resilience. Each pillar represents a convergence point where governance requirements identified through

empirical research align with established mechanisms from Airbus practice. By systematically addressing vulnerabilities and strengthening resilience capabilities, these guidelines facilitate the deployment of autonomous systems whose autonomy remains within the organization's demonstrated capacity for governance, monitoring, and rollback.

9

Discussion

9.1. Conditions for Selecting GenAI or Agentic AI

SQ1: *What data and system conditions justify choosing GenAI versus Agentic AI?*

The value of Agentic AI for supply chain resilience is contingent upon the presence of appropriate data and governance systems at specific decision points. The framework proposed in chapter 8 introduces two sequential evaluation layers at Supply Chain Operations Reference (SCOR) decision points. Initially, use cases are assessed against resilience guidelines (R1–R5), which evaluate vulnerability, capability balance, data quality, governance, and integration. Use cases that meet these criteria proceed to method-choice criteria (D1–D4), which examine orchestration requirements, the nature of work, risk-autonomy alignment, and organizational readiness.

The framework aligns artificial intelligence applications with Levels of Automation (LoA) as defined in Parasuraman's taxonomy (Parasuraman et al., 2000). Generative AI is appropriate when tasks require single interactions (D1) that depend on human judgment (D2), involve noisy or unstructured data (R3), and when systems permit information exposure but do not provide full execution control (D4). This configuration, characterized as "valuable but not governable," enables organizations to realize resilience benefits while preserving human oversight. Agentic AI requires multi-step orchestration (D1), execution-dominant repetitive tasks (D2), structured authoritative data (R3), and high readiness—including data maturity, robust interfaces, testing environments, monitoring, and rollback mechanisms (D4). Risk assessment (D3) calibrates autonomy: moderate, reversible consequences permit LoA 5–7; severe consequences mandate defaulting to manual control or GenAI advisory (LoA 1–4).

9.2. Safeguards and Human Oversight in Autonomous Operation

SQ2: *Which safeguards and human-in-the-loop controls are necessary to ensure safe and resilient autonomous operation?*

A comprehensive analysis of expert interviews, focus groups, and cross-industry analogies indicates that achieving safe and resilient autonomous operation requires more than model accuracy. Experts

consistently emphasize that hallucinations, non-deterministic behavior, and cascading multi-step errors—referred to in the interviews as “extreme agency”—are inherent to Agentic AI architectures rather than correctable defects. Therefore, safety relies on a robust system of governance, verification, and structural constraints integrated into the deployment architecture.

Empirical findings from both interview iterations and focus groups converge on eight transferable design patterns, which are detailed in Section 8.5. Collectively, these eight pillars address SQ2. Pillar I separates deliberation from execution through a layered control stack. Pillar II provides independent verification via the Command/Monitor pattern. Pillar III calibrates autonomy to the operational context through managed and selected modes. Pillar IV ensures safe behavior in the event of failure through graceful degradation. Pillar V enforces containment prior to diagnosis during disruption. Pillar VI prepares agents for out-of-distribution events using the Digital Supply Chain Twin. Pillar VII guarantees traceability and post-event learning through the Digital Black Box. Pillar VIII mitigates single-point failures through parallel architecture and redundant decision-making.

Together, these pillars establish a unified governance architecture in which human oversight is implemented as a formal supervisory control model rather than an informal backstop. Autonomy is calibrated to risk through Pillar III, independently verified through Pillars II and VIII, structurally separated by function through Pillar I, prepared for rare events through Pillar VI, safely degraded under failure through Pillars IV and V, and made auditable and recoverable through Pillar VII. Section 8.5 presents the full specification of each pillar, including the governance requirements addressed, the empirical basis for derivation, and the corresponding design mechanism.

9.3. Building Robustness to Unexpected Supply Chain Events

SQ3: *What training approaches improve an agent’s robustness to unexpected supply chain events?*

Resilience is determined by system performance during rare, out-of-distribution disruptions that extend beyond the operational envelope of historical training data. Assurance for such events must be established through systematic preparation rather than inferred from nominal performance. The second focus group and subsequent interviews identified three complementary strategies for embedding disruption readiness into Agentic AI systems.

First, synthetic scenario generation enables teams to construct extreme yet plausible disruption cases absent from historical records. This approach allows agents to rehearse retrieval, interpretation, justification, and escalation under stress without reliance on real-world crises.

Second, the Digital Supply Chain Twin offers a safe environment in which sequential multi-step behaviors can be evaluated and the downstream impact of early actions verified before execution privileges are granted, as detailed in the training capability mapping of Chapter 8.

Third, epistemic-aware methods enable agents to recognize when they are operating outside their reliable competence region and to implement safe degradation. This process involves transitioning from execution to recommendation-only mode, and ultimately to manual control, when uncertainty remains high or signals are conflicting.

To prevent unrealistic learning and fragile policies, synthetic scenarios must be validated by domain experts to ensure that counterfactual tests are plausible and consistent with actual operational constraints. Assurance processes should be evidence-based and auditable. Comprehensive logs of tool usage, de-

cisions, rationales, and anomaly triggers should constitute a Digital Black Box that facilitates incident review, accountability, and systematic updates to requirements and safeguards as new failure modes are identified. Disruption readiness is not a one-time setup but a continuous system design challenge, requiring detection mechanisms, conservative fallback strategies, and structured post-event learning as permanent features of the deployment architecture.

9.4. Aviation as External Validation of the Design Pillars

SQ4: *Which mission-critical aviation automation principles can be translated into practical design patterns for Agentic AI to improve supply chain resilience?*

The eight pillars presented in Section 8.5 are the result of a comprehensive empirical process conducted in this thesis. SQ4 specifically investigates whether structurally equivalent governance mechanisms exist in a mature, safety-critical automation domain, and whether such equivalence increases confidence in the applicability of these pillars.

Interviews with an Airbus line pilot and an Airbus AI and automation engineer demonstrate that each pillar has a direct structural precedent within the Airbus Automatic Flight System. Pillar I corresponds to the three-layer separation of Flight Management, Flight Guidance, and Fly-By-Wire. Pillar II aligns with the paired Flight Management and Guidance Computers, which cross-monitor and deactivate malfunctioning units to prevent error propagation. Pillar III is reflected in the FMGS dual-mode architecture, where crews transition from managed to selected control when predictability or data integrity is compromised. Pillar IV is exemplified by the progressive control law transition from Normal Law to Alternate Law and then to Direct Law. Pillar V is represented by emergency procedures that require stabilization before diagnosis. Pillar VI corresponds to the Iron Bird methodology and digital twin infrastructure, which are used to stress-test automation against synthetic scenarios prior to deployment. Pillar VII is reflected in the flight-recorder culture, which transforms operational anomalies into specification updates. Pillar VIII aligns with the triple-redundant air data and inertial reference units, which are cross-validated by majority voting, as well as the separated electrical and hydraulic systems.

In response to SQ4, aviation's contribution lies not in the origination of the pillars, which are derived from the empirical findings of this thesis, but in demonstrating that equivalent principles have been operationally validated in a high-reliability environment. This situates the pillars within a broader body of evidence that extends beyond the qualitative data collected for this research.

9.5. Designing and Governing Agentic AI for Supply Chain Resilience

RQ: *How should organizations design and govern Agentic AI systems to improve supply chain resilience?*

The four sub-questions addressed in the preceding sections converge on a single design and governance imperative: Agentic AI delivers value in supply chain resilience not through autonomy alone, but through autonomy that is explicitly bounded, systematically validated, and continuously supervised. SQ1 establishes that the choice between GenAI and Agentic AI depends on data maturity, orchestration requirements, risk-autonomy alignment, and organizational readiness at each SCOR decision point. Concepts must first demonstrate a credible resilience contribution before determining the level

and method of automation, ensuring that technological sophistication does not override resilience justification.

SQ2 demonstrates that safe autonomous operation requires the eight structural design patterns presented in Section 8.5. Collectively, these pillars implement human oversight as a designed supervisory control model rather than an informal backstop. They calibrate autonomy to risk, separate deliberation from execution, provide independent verification, ensure graceful degradation, and guarantee traceability.

SQ3 establishes that governance architecture alone is insufficient for the rare, high-impact events that define supply chain resilience. Agents must be systematically prepared through synthetic scenario generation, Digital Supply Chain Twin validation, and epistemic-aware methods that enable safe degradation when the system operates beyond its reliable competence region.

SQ4 confirms that the eight pillars are not context-specific heuristics derived from a limited qualitative sample, but are independently instantiated in aviation, a domain refined through decades of regulatory scrutiny and incident investigation. This provides external validity, extending the empirical grounding of the thesis beyond its qualitative data.

These findings yield three actionable recommendations for practitioners. First, managers should implement the eight pillars rather than pursue full end-to-end automation, situating AI capabilities where they provide the greatest resilience value and maintaining human oversight at critical decision points. Second, investment strategies should prioritize disruption readiness over exclusive focus on operational efficiency, allocating resources to stress-test agent performance in synthetic scenarios rather than relying solely on historical data. Third, effective deployment requires organizational preparedness that extends beyond technical integration, encompassing explicit protocols for human-AI collaboration, clearly defined escalation thresholds, and workforce competencies in both AI literacy and disruption management as prerequisites for responsible adoption.

10

Conclusion

The research methodology employed to address the research questions is a mixed-methods approach grounded in design thinking and implemented through a two-iteration Double Diamond (diverge-converge) process. This study integrates theoretical foundations with qualitative empirical evidence to link technical artificial intelligence (AI) capabilities to operational supply chain requirements during periods of disruption. The objective extends beyond collecting expert perspectives, aiming to translate these insights into a repeatable logic for system design and governance. The first iteration emphasizes exploration and co-design, primarily addressing SQ1 and SQ2. During the divergence stage, semi-structured interviews were conducted with three AI experts (representing consultancy, academia, and data science) and three supply chain experts (from industry, consultancy, and academia). These interviews clarified the practical boundary between GenAI and Agentic AI, and mapped the multi-layered risks associated with introducing autonomy into operational workflows. The interview data were analyzed using the six-step thematic analysis outlined by Braun and Clarke (2006), resulting in thematic-categories that describe capability boundaries, governance prerequisites, and recurring disruption propagation patterns. In the convergence stage, a focus group comprising a diverse set of experts translated these findings into actionable agent concepts linked to operational vulnerabilities. The group also established ”always allowed” and ”never allowed” boundaries, enabling governable autonomy to be articulated as operational constraints rather than general aspirations. To ensure consistent application of the first iteration’s outputs across use cases, the qualitative findings were formalized into an Integrated Framework for Resilience Evaluation and automation choice. This framework consists of two complementary steps. The first step screens whether a proposed idea plausibly enhances disruption detection, response coordination, or recovery. The second step evaluates the appropriate technological approach by assessing orchestration requirements, task characteristics, autonomy-risk alignment, and organizational readiness. This process provides a structured basis for determining whether a deterministic approach, GenAI, or Agentic AI is suitable for a specific use case and under which conditions it can be governed safely. The second iteration transitions from ideation to validation and robustness refinement, addressing SQ3 and SQ4. A focused aviation case study was conducted to extract transferable principles from a high-reliability domain, utilizing interviews with an Airbus line pilot and an Airbus AI/automation engineer. These interviews facilitated the translation of safety-oriented design

patterns into implications for Agentic system governance. Concurrently, targeted validation interviews were performed to stress-test the initial agent concepts and refine both the guidelines and deployment conditions within the integrated framework. A subsequent focus group examined disruption readiness, evaluating methods to prepare Agentic AI for out-of-distribution events, including synthetic scenario creation, digital supply chain twins (DSCT), and epistemic-aware methods. To enhance reliability and minimize bias, the study implemented a structured data preparation and quality control protocol. Interviews underwent a three-stage transcription process: automated transcription (Google Notebook LM), manual correction (Microsoft Word), and structural refinement (Headstart). In summary, the methodology addresses the research questions by employing semi-structured interviews and focus group to establish the initial logic for method selection and safeguards (SQ1, SQ2 and SQ3), and by utilizing cross-domain analysis for SQ4. This structured synthesis enables the central research question on design and governance to be addressed, translating empirical insights into explicit decision rules for the safe and effective deployment of varying levels of autonomy to enhance supply chain resilience.

10.1. Research Limitations

This research utilizes interviews with AI, supply chain, and Airbus experts, as well as two focus groups. Although this approach provides valuable contextual insights, the limited sample size restricts statistical generalizability. Additionally, the majority of participants were affiliated with European organizations, specifically from the Netherlands, Italy, Germany, and France, and represented digitally mature organizations in the consumer goods, aerospace, and consulting sectors. This geographic and sectoral concentration may introduce selection bias and limit the diversity of perspectives, thereby constraining the applicability of findings to firms at earlier stages of digital transformation or those operating in different regional contexts. The cross-sectional design captures perspectives at a single point in time amid rapid technological change, which prevents observation of how Agentic AI implementations and governance structures evolve or how initial design principles perform under changing conditions. This temporal limitation is particularly significant because both Agentic AI technology and its regulatory environment were evolving during the research period, and subsequent developments may require recalibration of specific guidance. Another limitation stems from the conceptual nature of the proposed frameworks and agent portfolios. These constructs were developed through expert consultation and focus groups but have not been validated in operational environments. As a result, critical questions remain unresolved, including actual performance under disruption scenarios, practical feasibility of governance mechanisms, and genuine cost–benefit trade-offs. Moreover, SCOR was selected as the reference language due to its established process architecture and metric library; however, alternative reference models could have offered complementary perspectives. Technical limitations of current Agentic AI, including hallucination risks, context window constraints, and computational costs, render certain recommendations provisional. In addition, the prevailing opacity of AI decision-making, as most large language and agentic models provide limited or no interpretable justification for their outputs, raises further concerns for governance and accountability within supply chain contexts. The rapid pace of AI development and the evolving regulatory landscape also introduce uncertainty regarding the longevity of specific guidance. Notably, several model generations and capability milestones emerged during the course of this research, indicating that some technical assessments may already reflect an outdated landscape. Qualitative analysis necessarily involves researcher interpretation, and the analogy between aviation and supply chain domains assumes a level of comparability that may not fully apply. Aviation systems benefit from decades of standardization, mandatory certification, and deterministic

architectures, which differ significantly from the heterogeneous and commercially negotiated nature of global supply chains. Therefore, the Airbus-derived design principles should be regarded as heuristics that require contextual adaptation. These limitations delineate the appropriate scope of claims and underscore opportunities for future research.

10.2. Future Research

The limitations identified in this thesis indicate several priorities for future research. Foremost, empirical validation studies are needed to test the proposed frameworks and agent portfolios in real-world environments. Controlled pilot implementations that compare organizations with and without Agentic AI capabilities could address unresolved questions regarding performance during disruptions, the practical feasibility of governance mechanisms, and actual cost–benefit trade-offs. Longitudinal tracking would capture the evolution of agent portfolios, governance structures, and human–AI collaboration as both technology and user competencies develop, thereby directly addressing the temporal constraints inherent in this cross-sectional design. Future research should expand sample diversity beyond digitally mature European organizations. Studies involving firms at earlier stages of digital transformation, small and medium-sized enterprises, and organizations in emerging markets would test the generalizability of these findings across varying resource constraints and institutional contexts. Expanding the sectoral focus to include healthcare supply chains, humanitarian logistics, and service industries would help determine which design principles are robust and which require adaptation to specific contexts. The exclusive reliance on SCOR as the reference framework highlights opportunities for comparative analysis using alternative models, which may identify different optimal intervention points for Agentic AI. Additionally, the analogy between aviation and supply chains requires systematic examination to determine its validity and limitations. Research that draws on other safety-critical domains could either reinforce confidence in the Airbus-derived pillars or indicate necessary modifications for application within supply chains. Technical research should address current limitations of Agentic AI by advancing out-of-distribution detection, explainability mechanisms, and strategies for graceful degradation. Simulation-based testing in digital twin environments would facilitate systematic exploration of agent behavior under synthetic disruptions prior to operational deployment. As the regulatory landscape evolves, particularly with the implementation of the EU AI Act, research should monitor and adapt governance frameworks accordingly. Pursuing these directions will enable a transition from exploratory frameworks to empirically validated approaches that organizations can deploy with greater confidence.

10.3. Managerial Implications and Societal Relevance

This research offers supply chain managers practical guidance for implementing Agentic AI. The findings refute the assumption that autonomy alone ensures resilience, instead demonstrating that value arises from intentional architectural decisions that balance autonomy and control. Three principal implications are identified. First, managers are advised to implement a layered control architecture instead of full end-to-end automation, situating AI capabilities where they generate the greatest value and retaining human oversight at key decision points. Second, investment strategies should emphasize disruption readiness rather than focusing solely on operational efficiency, directing resources toward evaluating agent performance in novel scenarios rather than relying only on historical data. Third, effective deployment necessitates organizational preparedness that extends beyond technical integration, encompassing explicit protocols for human–AI collaboration, clearly defined escalation thresholds, and

workforce competencies in both AI literacy and disruption management. In addition to firm-level advantages, this research addresses wider societal challenges. Supply chains represent critical infrastructure that supports economic stability, public health, and access to essential goods. As disruptions become more frequent, the ability of these networks to absorb shocks without widespread failures constitutes a public good. The design principles presented in this study enhance network robustness and enable organizations to respond more effectively to environmental and regulatory demands. This work also examines issues related to AI governance and workforce adaptation. As Agentic AI acquires increased decision-making authority, there is a societal need for frameworks that ensure safety, transparency, and human oversight. The proposed design principles delineate boundaries that mitigate over-reliance on automation while maintaining meaningful human agency. The transition toward human–AI collaboration requires workforce development, as roles shift from routine tasks to higher-level judgment and system supervision. Collaboration among policymakers, industry, and educational institutions is essential to ensure that Agentic AI deployment reinforces, rather than undermines, societal cohesion.

References

Ahmed, T., Karmaker, C. L., Nasir, S. B., Moktadir, M. A., & Paul, S. K. (2023). Modeling the artificial intelligence-based imperatives of industry 5.0 towards resilient supply chains: A post-COVID-19 pandemic perspective. *Computers & Industrial Engineering*, *177*, 109055. https://doi.org/10.1016/j.cie.2023.109055

Airbus. (1998). *A319/a320/a321 flight deck & systems briefing for pilots*. Airbus Flight Operations Support. Toulouse, France.

Airbus. (2008). *22 auto flight system presentation: A330-200/300 Technical Training Manual*. https://www.scribd.com/document/471438995/ATA-22-pdf

Airbus. (2017). *Taking flight with the airbus "iron bird"* [Web Story (Innovation), published 16 May 2017]. https://www.airbus.com/en/newsroom/news/2017-05-taking-flight-with-the-airbus-iron-bird

Airbus. (2018, March). Airbus developed a350 xwb safety feature enables automated emergency descents.

Airbus. (2022). *Virtual flight-test campaign and lab test paving way for a321xlr's first flight* [Web Story (Commercial Aircraft), published 31 May 2022]. https://www.airbus.com/en/newsroom/stories/2022-05-virtual-flight-test-campaign-and-lab-test-paving-way-for-a321xlrs-first

Airbus. (2025a). *Digital twins: Accelerating aerospace innovation from design to operations* [Web Story (Innovation), published 24 April 2025]. https://www.airbus.com/en/newsroom/stories/2025-04-digital-twins-accelerating-aerospace-innovation-from-design-to-operations

Airbus. (2025b). *Further preventing loss of control in-flight*. https://safetyfirst.airbus.com/further-preventing-loss-of-control-in-flight/

Airbus. (2025c). *What is a full flight simulator?* [Web Story (Commercial Aircraft), published 26 September 2025]. https://www.airbus.com/en/newsroom/stories/2025-09-what-is-a-full-flight-simulator

Airbus, P. S. D. (2019, July). Safety first #28 — airbus safety magazine, july 2019 [© Airbus S.A.S. 2019.].

Akter, S., Bandara, R., Hossain, M. N., Wamba, S. F., Foropon, C., & Papadopoulos, T. (2021). Analytics-enabled business process management: A review and framework. *Journal of Business Research*, *124*, 335–348.

APICS/ASCM. (2017). *SCOR® version 12.0: The framework—introduction* [Chicago, IL: APICS. SCOR processes: Plan, Source, Make, Deliver, Return, Enable; Levels 1–3]. APICS.

ASCM. (n.d.). The ABCs of supply chain resilience.

Association for Supply Chain Management. (2025). *Top 10 supply chain trends 2025*. ASCM.

Australian Transport Safety Bureau. (2018, January). *Descent below segment minimum safe altitude during a non-precision instrument approach involving airbus a320, pk-axy, 17 km wsw perth airport, western australia, 19 february 2016* (Transport Safety Report No. AO-2016-012 (Final)) (© Commonwealth of Australia 2018). Australian Transport Safety Bureau. https://www.atsb.gov.au/sites/default/files/media/5773855/ao-2016-012_final.pdf

Aydemir, H., Zengin, U., Fischer, O. J. P., Durak, U., & Hartmann, S. (2025). From iron birds to digital twins with engineering simulators: Toward virtual certification. *Journal of Aerospace Information Systems*, 1–10.

Balaji, M., Dinesh, S. N., Kumar, P. M., & Hari Ram, K. (2021). Balanced scorecard approach in deducing supply chain performance. *Materials Today: Proceedings*, *47*, 5217–5222.

Balbix. (n.d.). *Understanding agentic AI and its cybersecurity applications*. https://www.balbix.com/insights/understanding-agentic-ai-and-its-cybersecurity-applications/

Barman, D., Saha, S., & Pal, D. (2024). The dark side of language models. *Patterns*, *5*(7), 100986. https://doi.org/10.1016/j.patter.2024.100986

Behzadi, G., O'Sullivan, M. J., & Olsen, T. L. (2020). On metrics for supply chain resilience. *International Journal of Production Economics*, *230*, 107793. https://doi.org/10.1016/j.ijpe.2020.107793

Belhadi, A., Kamble, S., Fosso Wamba, S., & Queiroz, M. M. (2022). Building supply-chain resilience: An artificial intelligence-based technique and decision-making framework. *International Journal of Production Research*, *60*(14), 4487–4507. https://doi.org/10.1080/00207543.2021.1950935

Belhadi, A., Mani, V., Kamble, S. S., & et al. (2024). Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: An empirical investigation. *Annals of Operations Research*, *333*, 627–652. https://doi.org/10.1007/s10479-021-03956-x

Bender, E. M., & Koller, A. (2020). Climbing towards nlu: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. https://doi.org/10.18653/v1/2020.acl-main.463

Beta, K., Nagaraj, S. S., & Weerasinghe, T. D. B. (2025). The role of artificial intelligence on supply chain resilience. *Journal of Enterprise Information Management*, *38*(3), 950–973. https://doi.org/10.1108/JEIM-12-2023-0674

Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2001). *Focus groups in social research*. SAGE Publications Ltd. https://doi.org/10.4135/9781849209175

Boone, T., Fahimnia, B., Ganeshan, R., Herold, D. M., & Sanders, N. R. (2025). Generative ai: Opportunities, challenges, and research directions for supply chain resilience. *Transportation Research Part E: Logistics and Transportation Review*, *199*, 104135.

Boston Consulting Group. (2025). Cost and resilience: The new supply chain challenge. https://www.bcg.com/publications/2025/cost-resilience-new-supply-chain-challenge

Boyes, H., & Watson, T. (2022). Digital twins: An analysis framework and open issues. *Computers in Industry*, *143*, 103763.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Brière, D., Favre, C., & Traverse, P. (1995). A family of fault-tolerant systems: Electrical flight controls, from airbus a320/330/340 to future military transport aircraft. *Microprocessors and Microsystems*, *19*(2), 75–82. https://doi.org/https://doi.org/10.1016/0141-9331(95)98982-P

Brière, D., Favre, C., & Traverse, P. (2001). Electrical flight controls, from airbus a320/330/340 to future military transport aircraft: A family of fault-tolerant systems. In *The avionics handbook*. CRC Press.

Bruneau, M., Chang, S., Eguchi, R., Lee, G., O'Rourke, T., Reinhorn, A., Shinozuka, M., Tierney, K., Wallace, W., & von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra*, *19*(4), 733–752.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Cao, C., Wang, F., Lindley, L., & Wang, Z. (2024). Managing Linux servers with LLM-based AI agents: An empirical evaluation with GPT4. *Machine Learning with Applications*, *17*, 100570. https://doi.org/10.1016/j.mlwa.2024.100570

Chigot, E., Wilson, D. G., Ghrib, M., Jimenez, F., & Oberlin, T. (2025). Synthetic data for robust runway detection. *International Conference on Computer Analysis of Images and Patterns*, 294–304.

Chowdhury, M. M. H., & Quaddus, M. (2017a). Supply chain readiness, response and recovery for resilience. *Supply Chain Management: An International Journal*, *22*(2), 116–138.

Chowdhury, M. M. H., & Quaddus, M. (2017b). Supply chain resilience: Conceptualization and scale development using dynamic capability theory. *International Journal of Production Economics*, *188*, 185–204. https://doi.org/10.1016/j.ijpe.2017.03.020

Christensen, B. T., & Schunn, C. D. (2007). The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design. *Memory & cognition*, *35*(1), 29–38.

Christopher, M., & Peck, H. (2004). Building the resilient supply chain. *The International Journal of Logistics Management*, *15*(2), 1–14.

Chukwu, N., Sevidzem Simo, Y., Ejiofor, O., Ekweli, D., et al. (2024). Resilient chain: AI-enhanced supply chain security and efficiency integration. *International Journal of Scientific and Management Research*, *7*(3), 46–65. https://doi.org/10.37502/IJSMR.2024.7306

Corrao, S. (2002). *Il focus group* (2nd ed.). FrancoAngeli.

Council, D. (2025). Framework for innovation. https://www.designcouncil.org.uk/our-resources/framework-for-innovation/

Crabtree, B. F., Yanoshik, M. K., Miller, W. L., & O'Connor, P. J. (1993). Selecting individual or group interviews. In D. L. Morgan (Ed.), *Successful focus groups: Advancing the state of the art* (pp. 137–149). Sage.

Culot, G., Orzes, G., Sartor, M., & Nassimbeni, G. (2024). Artificial intelligence in supply chain management: A systematic literature review of empirical studies. *International Journal of Production Economics*, *269*, 109865. https://doi.org/10.1016/j.ijpe.2024.109865

Delipinar, G. E., & Kocaoglu, B. (2016). Using scor model to gain competitive advantage: A literature review. *Procedia-Social and Behavioral Sciences*, *229*, 398–406.

Deloitte & ASCM. (2020). Digital capabilities model for supply chain [A practitioner framework building upon SCOR for digital supply chain transformation].

Deloitte AI Institute. (2025). *The state of generative ai in the enterprise 2024: Year-end generative ai report* [Landing page for the Q1–Q4 series]. Deloitte. https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-generative-ai-in-enterprise.html

Deloitte Insights. (2024). Supply chain resilience amid disruptions. https://www.deloitte.com/us/en/insights/industry/manufacturing-industrial-products/global-supply-chain-resilience-amid-disruptions.html

Deloitte Insights. (2025). Enhancing supply chain resilience in a new era of policy: Managing supply chains amid tariffs. https://www.deloitte.com/us/en/insights/industry/manufacturing/managing-supply-chains-amid-tariffs.html

Deloitte UK. (2022). The imperative for increased supply chain resilience. https://www.deloitte.com/uk/en/services/consulting-risk/research/the-imperative-for-increased-supply-chain-resilience.html

d'Enquêtes et d'Analyses, B. (2012). *Final report: Accident on 1 june 2009 to the airbus a330-203, air france 447*. BEA. Paris, France.

DHL & Analytics, E. (2025, January). *Top 5 supply chain risks in 2025*. https://www.dhl.com/global-en/delivered/global-trade/top-5-supply-chain-risks-in-2025.html

Diaferia, L., Blohm, I., De Rossi, L. M., Salviotti, G., et al. (2022). When standard is not enough: A conceptualization of ai systems' customization and its antecedents. In *Icis 2022 proceedings*.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Williams, M. D., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy [https://doi.org/10.1016/j.ijinfomgt.2019.08.002]. *Journal of Business Research*, *122*, 725–748.

Emenike, L. (2025). *The difference between genai, ai agents, and agentic AI*. https://lawrence-emenike.medium.com/the-difference-between-genai-ai-agents-and-agentic-ai-5b66fbb462b0

Enkel, E., & Gassmann, O. (2010). Creative imitation: Exploring the case of cross-industry innovation. *R&d Management*, *40*(3), 256–270.

European Commission, DG CONNECT. (2025). AI Act — shaping europe's digital future [Application timeline and GPAI guidance; page last updated 1 Aug 2025].

Falasca, M., Zobel, C., & Cook, D. (2008). A decision support framework to assess supply chain resilience. *Proceedings of the 5th International ISCRAM Conference*, 596–605.

Fan, T., Lin, X., Fu, C., Yeung, J. H.-y., & Shi, C. (2023). Managing supply chain disruptions: Responding to the impacts of covid-19 on supply chains. *International Journal of Production Economics*, *255*, 108670. https://doi.org/10.1016/j.ijpe.2022.108670

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S., & Li, Q. (2024). A survey on rag meeting llms: Towards retrieval-augmented large language models. *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 6491–6501.

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, *630*, 625–630. https://doi.org/10.1038/s41586-024-07421-0

Federal Aviation Administration. (2022). *Optimum use of automation — flight operations briefing notes (flt_ops-sop-seq02)* (Flight Operations Briefing Notes No. FLT_OPS-SOP-SEQ02). Federal Aviation Administration. https://www.faa.gov/sites/faa.gov/files/2022-11/AirbusSafetyLib_-FLT_OPS-SOP-SEQ02%20-%20Automation.pdf

Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., & You, J. (2024). How far are we from agi: Are llms all we need? *arXiv preprint arXiv:2405.10313*.

First, A. S. (2021, September). Safe handling of tcas alerts [© Airbus Safety First].

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, *2*(1).

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy [doi: 10.1016/S0364-0213(83)80009-3]. *Cognitive Science*, *7*(2), 155–170.

Gerger Swartling, Å. (2007). *Focus groups* (Guidance chapter). Stockholm Environment Institute.

Goel, A. K. (2002). Design, analogy, and creativity. *IEEE expert*, *12*(3), 62–70.

Goteman, Ö. E., & Dekker, S. (2006). Flight crew callouts and aircraft automation modes. *The International Journal of Applied Aviation Studies*, *6*(2), 291–304.

Goupil, P. (2011a). Airbus state of the art and practices on fdi and ftc in flight control system [SAFEPROCESS 2009]. *Control Engineering Practice*, *19*(6), 524–539. https://doi.org/https://doi.org/10.1016/j.conengprac.2010.12.009

Goupil, P. (2011b). Airbus state of the art and practices on fdi/ftc in flight control systems. *Annual Reviews in Control*, *35*(2), 171–182. https://doi.org/10.1016/j.arcontrol.2011.03.008

Gunasekaran, A., Subramanian, N., & Rahman, S. (2015). Supply chain resilience: Role of complexities and strategies. *International Journal of Production Research*, *53*(22), 6809–6819. https://doi.org/10.1080/00207543.2015.1093667

Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, *38*, 100285. https://doi.org/10.1016/j.cosrev.2020.100285

Hendricks, K. B., & Singhal, V. R. (2005). An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations Management*, *14*(1), 35–52. https://doi.org/10.1111/j.1937-5956.2005.tb00008.x

Hill, C. A., Zhang, G. P., & Miller, K. E. (2018). Collaborative planning, forecasting, and replenishment & firm performance: An empirical evaluation. *International Journal of Production Economics*, *196*, 12–23. https://doi.org/10.1016/j.ijpe.2017.11.012

Hohenstein, N.-O., Feisel, E., Hartmann, E., & Giunipero, L. (2015). Research on the phenomenon of supply chain resilience: A systematic review and paths for further investigation. *International Journal of Physical Distribution & Logistics Management*, *45*(1/2), 90–117. https://doi.org/10.1108/IJPDLM-05-2013-0128

Holland, M., & Chaudhari, K. (2024). Large language model based agent for process planning of fiber composite structures. *Manufacturing Letters*, *40*, 100–103. https://doi.org/10.1016/j.mfglet.2024.03.010

Hosseini, S., Ivanov, D., & Dolgui, A. (2019). Review of quantitative methods for supply chain resilience analysis. *Transportation Research Part E: Logistics and Transportation Review*, *125*, 285–307. https://doi.org/10.1016/j.tre.2019.03.001

Huang, S. H., Sheoran, S. K., & Keskar, H. (2005). Computer-assisted supply chain configuration based on SCOR. *Computers & Industrial Engineering*, *48*(2), 377–394. https://doi.org/10.1016/j.cie.2005.01.001

IBM. (n.d.-a). *Agentic AI vs. generative AI*. https://www.ibm.com/think/topics/agentic-ai-vs-generative-ai

IBM. (n.d.-b). What is supply chain resilience? https://www.ibm.com/think/topics/supply-chain-resiliency

Ivanov, D. (2021a). Disruption tails and revival policies: A simulation analysis of supply chain design and production-ordering systems in the recovery and post-disruption periods. *Computers & Industrial Engineering*, *158*, 107444. https://doi.org/10.1016/j.cie.2021.107444

Ivanov, D. (2021b). *Introduction to supply chain resilience: Management, modelling, technology*. Springer Nature.

Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: Extending the supply chain resilience angles towards survivability. *International Journal of Production Research*, *58*(10), 2904–2915. https://doi.org/10.1080/00207543.2020.1750727

Ivanov, D., & Sokolov, B. (2010). *Adaptive supply chain management*. Springer.

Jain, V., Kumar, S., Soni, U., & Chandra, C. (2017). Supply chain resilience: Model development and empirical analysis. *International Journal of Production Research*, *55*(22), 6779–6800. https://doi.org/10.1080/00207543.2017.1349947

Kamalahmadi, M., & Parast, M. M. (2016). A review of the literature on the principles of enterprise and supply chain resilience: Major findings and directions for future research. *International Journal of Production Economics*, *171*(1), 116–133. https://doi.org/10.1016/j.ijpe.2015.10.023

Kilic, U., Cam, O., & Can, E. (2024). Machine learning–aided synthetic air data system for commercial aircraft. *Journal of Aerospace Engineering*, *37*(6), 04024071.

Kocaoğlu, B., Gülsün, B., & Tanyas, M. (2013). A scor based approach for measuring a benchmarkable supply chain performance. *Journal of Intelligent Manufacturing*, *24*(1), 113–132. https://doi.org/https://doi.org/10.1007/s10845-011-0547-z

Kontio, J., Bragge, J., & Lehtola, L. (2008). The focus group method as an empirical tool in software engineering. In F. Shull, J. Singer, & D. I. K. Sjøberg (Eds.), *Guide to advanced empirical software engineering* (pp. 93–116). Springer. https://doi.org/10.1007/978-1-84800-044-5_4

Kumawat, P. (2024). Impact of artificial intelligence in building supply chain resiliency. *International Journal of Supply Chain Management*, *13*(6), 10–20. https://doi.org/10.59160/ijscm.v13i6.6283

Lambert, D. M., & Cooper, M. C. (2000). Issues in supply chain management. *Industrial Marketing Management*, *29*(1), 65–83. https://doi.org/10.1016/S0019-8501(99)00113-3

Lambert, D. M., & Enz, M. G. (2017). Issues in supply chain management: Progress and potential. *Industrial Marketing Management*, *62*, 1–16. https://doi.org/10.1016/j.indmarman.2016.12.002

Lelaie, C. (2012, January). *A380: Development of the flight controls — part 1* [Published in *Safety First* #13, © Airbus Safety First]. https://safetyfirst.airbus.com/app/themes/mh_newsdesk/documents/archives/a380-development-of-the-flight-controls2.pdf

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459–9474.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 9459–9474. https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

Lim, Weng, & Marc. (2024). What is qualitative research? an overview and guidelines. *Australasian Marketing Journal*, *33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Lim, W. M. (2025). What is qualitative research? an overview and guidelines. *Australasian Marketing Journal*, *33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Lima-Junior, F. R., & Carpinetti, L. C. R. (2019). Predicting supply chain performance based on SCOR® metrics and multilayer perceptron neural networks. *International Journal of Production Economics*, *212*, 19–38. https://doi.org/10.1016/j.ijpe.2019.02.001

Liu, J. (2017, September 14). *Visualizing the 4 essentials of design thinking*. Medium. https://medium.com/good-design/visualizing-the-4-essentials-of-design-thinking-17fe5c191c22

Lockamy, A., & McCormack, K. (2004). The development of a supply chain management process maturity model using the SCOR model. *Supply Chain Management: An International Journal*, *9*(4), 272–278. https://doi.org/10.1108/13598540410550019

Malinge, Y. (2016). Understanding aoa margins and vαprot on protected aircraft. *Safety First*, *21*, 18–27.

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., … Oak, S. (2025). *Artificial intelligence index report 2025* (Also

available as arXiv:2504.07139). Stanford Institute for Human-Centered Artificial Intelligence (HAI). https://doi.org/10.48550/arXiv.2504.07139

McCarthy, J. (2007). *What is artificial intelligence?* http://www-formal.stanford.edu/jmc/

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12–12.

McKinsey & Company. (2025). *The state of AI: How organizations are rewiring to capture value*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Meafa, A.-E., Chaouni Benabdellah, A., & Zekhnini, K. (2024). Enhancing supply chain resilience through dynamic capabilities of blockchain technology: A structural model analysis. *Procedia Computer Science*, *232*, 980–989. https://doi.org/10.1016/j.procs.2024.01.097

Medetalibeyoglu, A., Velichko, Y. S., Hart, E. M., & Bagci, U. (2024). Foundational artificial intelligence models and modern medical practice. *BJR|Artificial Intelligence*, *2*(1), ubae018. https://doi.org/10.1093/bjrai/ubae018

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology*, *51*(6), 541–557. https://www.jstor.org/stable/2770681

Miltner, M., Duan, P. P., & de Haag, M. U. (2014). Modeling and utilization of synthetic data for improved automation and human-machine interface continuity. *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*, 2D4–1.

Moreira, N. (2020, February 13). *A diverge–converge technique to solve a problem*. Medium. https://nmoreira.medium.com/a-diverge-converge-technique-to-solve-a-problem-1f8b7d78553c

Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology*, *22*, 129–152. https://doi.org/10.1146/annurev.soc.22.1.129

National Institute of Standards and Technology. (2024a). *Artificial intelligence risk management framework: Generative ai profile (ai 600-1)*. https://doi.org/10.6028/NIST.AI.600-1

National Institute of Standards and Technology. (2024b). *Generative ai profile (companion to the ai risk management framework)* (tech. rep. No. NIST AI 600-1). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

National Institute of Standards and Technology. (2024c, July). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (tech. rep. No. NIST AI 600-1). NIST. Gaithersburg, MD. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

Naz, F., Kumar, A., Majumdar, A., & Agrawal, R. (2022). Is artificial intelligence an enabler of supply chain resiliency post COVID-19? an exploratory state-of-the-art review for future research. *Operations Management Research*, *15*, 378–398. https://doi.org/10.1007/s12063-021-00208-w

Niu, Y., Werle, N., Cohen, M., Cui, S., Deshpande, V., Ernst, R., Huchzermeier, A., Tsay, A. A., & Wu, J. (2025). Restructuring global supply chains: Navigating challenges of the covid-19 pandemic and beyond. *Manufacturing & Service Operations Management*, *27*(4), 1025–1036. https://doi.org/10.1287/msom.2024.0879

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Ntabe, E. N., LeBel, L., Munson, A. D., & Santa-Eulalia, L. A. (2015). A systematic literature review of the supply chain operations reference (SCOR) model application with emphasis on green supply chain management. *International Journal of Production Economics*, *169*, 310–332. https://doi.org/10.1016/j.ijpe.2015.08.008

OECD. (2025). *Oecd supply chain resilience review: Navigating risks*. OECD Publishing. https://doi.org/10.1787/94e3a8ea-en

Olsen, W. K. (2004). *Triangulation in social research: Qualitative and quantitative methods can really be mixed* [Editors: Haralambos, M. and Holborn, M. (Eds.)]. Causeway Press Ltd.

Omar, M., Sorin, V., Collins, J. D., Reich, D., Freeman, R., Gavin, N., Charney, A., Stump, L., Bragazzi, N. L., Nadkarni, G. N., & Klang, E. (2025). Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine*, *5*(330). https://doi.org/10.1038/s43856-025-01021-3

Palma-Mendoza, J. A. (2014). Analytical hierarchy process and SCOR model to support supply chain re-design. *International Journal of Information Management*, *34*(5), 634–638. https://doi.org/10.1016/j.ijinfomgt.2014.06.002

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286–297.

Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research & Method in Education*, *29*(1), 23–37. https://doi.org/10.1080/01406720500537304

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice* (4th). Sage.

Payne, T. (2025). Build supply chain resilience to arrive at an antifragile state. https://www.gartner.com/en/articles/supply-chain-resilience

Persson, F. (2011). SCOR template—a simulation based dynamic supply chain analysis tool. *International Journal of Production Economics*, *131*(1), 288–294. https://doi.org/10.1016/j.ijpe.2010.09.029

Pettit, T. J., Croxton, K. L., & Fiksel, J. (2019). The evolution of resilience in supply chain management: A retrospective on ensuring supply chain resilience. *Journal of Business Logistics*, *40*(1), 56–65. https://doi.org/10.1111/jbl.12202

Pettit, T. J., Fiksel, J., & Croxton, K. L. (2010). Ensuring supply chain resilience: Development of a conceptual framework. *Journal of Business Logistics*, *31*(1), 1–21. https://doi.org/10.1002/j.2158-1592.2010.tb00125.x

Piccialli, F., Chiaro, D., Sarwar, S., Cerciello, D., Qi, P., & Mele, V. (2025). Agentai: A comprehensive survey on autonomous agents in distributed ai for industry 4.0. *Expert Systems with Applications*, *291*, 128404. https://doi.org/10.1016/j.eswa.2025.128404

Pogna, A., Zhang, K., Zuiderwijk, A., & Stokkink, P. (2025). *Assessing the integration of freight transportation into public transport: A stakeholder analysis of barriers and opportunities* (tech. rep.). Center for Open Science.

Ponomarov, S. Y., & Holcomb, M. C. (2009). Understanding the concept of supply chain resilience. *The International Journal of Logistics Management*, *20*(1), 124–143. https://doi.org/10.1108/09574090910954873

Ren, Y., Liu, Y., Ji, T., & Xu, X. (2025). Ai agents and agentic ai—navigating a plethora of concepts for future manufacturing. *Journal of Manufacturing Systems*, *83*, 126–133. https://doi.org/10.1016/j.jmsy.2025.06.012

Riad, M., Naimi, M., & Okar, C. (2024). Enhancing supply chain resilience through artificial intelligence: Developing a comprehensive conceptual framework. *Logistics*, *8*(4), 111. https://doi.org/10.3390/logistics8040111

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson. https://aima.cs.berkeley.edu/4th-ed/pdfs/newchap02.pdf

Russell, S., Norvig, P., & Intelligence, A. (1995). A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, *25*(27), 79–80.

SAE International. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles* (J3016_202104).

SAP. (2025a). *Moving from genai to agentic AI within the customer experience*. https://www.forbes.com/sites/sap/2025/07/23/moving-from-genai-to-agentic-ai-within-the-customer-experience/

SAP. (2025b, January). *Supply chain predictions and outlook for 2025: Forbes brandvoice*. https://www.forbes.com/sites/sap/2025/01/16/supply-chain-predictions-and-outlook-for-2025/

SAP. (n.d.). What is a resilient supply chain? https://www.sap.com/products/scm/integrated-business-planning/what-is-a-resilient-supply-chain.html

Sapkota, R., Rivas, P., Roy, S., & Ghasem-Aghaee, N. (2025). Ai agents vs. agentic ai: A conceptual taxonomy, principles, challenges, and opportunities. *Information Fusion*, *120*, 102431. https://doi.org/10.1016/j.inffus.2025.102431

Sari, K. (2008). On the benefits of CPFR and VMI: A comparative simulation study. *International Journal of Production Economics*, *113*(2), 575–586. https://doi.org/10.1016/j.ijpe.2007.10.021

Sheffi, Y., & Rice, J. B. (2005). A supply chain view of the resilient enterprise. *MIT Sloan Management Review*, *47*(1), 41–48.

Simchi-Levi, D., Schmidt, W., & Wei, Y. (2014). From superstorms to factory fires: Managing unpredictable supply-chain disruptions. *Harvard Business Review*, *92*(1-2), 96–101.

Simchi-Levi, D., Wang, H., & Wei, Y. (2018). Increasing supply chain robustness through process flexibility and inventory. *Production and Operations Management*, *27*(8), 1476–1491.

Sinfield, J. V., Gustafson, T., & Hindo, B. (2013). The discipline of creativity. *MIT Sloan Management Review*, *55*(2), 24–26. https://sloanreview.mit.edu/article/the-discipline-of-creativity/

Singh, C. S., Soni, G., & Badhotiya, G. K. (2019). Performance indicators for supply chain resilience: Review and conceptual framework. *Journal of Industrial Engineering International*, *15*(Suppl 1), 105–117. https://doi.org/10.1007/s40092-019-00322-2

Singh, D., Sharma, A., Singh, R. K., & Rana, P. S. (2025). Augmenting supply chain resilience through AI and big data. *Business Process Management Journal*, *31*(2), 631–657. https://doi.org/10.1108/BPMJ-04-2024-0260

SKYbrary. (2025). *Flight control laws* [Copyright © SKYbrary Aviation Safety]. https://skybrary.aero/articles/flight-control-laws

S&P Global Market Intelligence. (2024). *2025 supply chain outlook: Dealing with consequences*. S&P Global.

Sprockhoff, J., Gupta, S., Durak, U., & Krueger, T. (2024). Scenario-based synthetic data generation for an ai-based system using a flight simulator. *AIAA SciTech 2024 Forum*, 1462.

Stefani, T., Christensen, J. M., Girija, A. A., Gupta, S., Durak, U., Köster, F., Krüger, T., & Hallerbach, S. (2025). Automated scenario generation from operational design domain model for testing ai-based systems in aviation. *CEAS Aeronautical Journal*, *16*(1), 197–212.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice*. Sage.

Stewart, G. (1997). Supply-chain operations reference (SCOR): The first cross-industry framework for integrated supply-chain management. *Logistics Information Management*, *10*(2), 62–67. https://doi.org/10.1108/09576059710815716

Strelkova, O. (2017). Three types of artificial intelligence [Conference paper, Khmelnitsky National University, Ukraine]. *Current Trends in Young Scientists' Research*, 1–4.

Svensson, C., & Hvolby, H.-H. (2012). Establishing a business process reference model for universities. *Procedia Technology*, *5*, 635–642. https://doi.org/10.1016/j.protcy.2012.09.070

Swartling, Å. G. (2007). Focus groups [European Commission webbook chapter]. In *Advanced tools for sustainability assessment: European commission webbook*. IVM web site. http://www.ivm.vu.nl/en/projects/Archive/SustainabilityA-test/index.asp

Tawse, A., & Tabesh, P. (2023). Thirty years with the balanced scorecard: What we have learned. *Business Horizons*, *66*(1), 123–132. https://doi.org/10.1016/j.bushor.2022.03.005

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research*, *122*, 502–517. https://doi.org/10.1016/j.jbusres.2020.09.009

Traverse, P., Lacaze, I., & Souyris, J. (2006). Airbus fly-by-wire: A process toward total dependability. *Proceedings of the 25th International Congress of the Aeronautical Sciences (ICAS)*.

Tukamuhabwa, B. R., Stevenson, M., Busby, J., & Zorzini, M. (2015). Supply chain resilience: Definition, review and theoretical foundations for further study. *International Journal of Production Research*, *53*(18), 5592–5623. https://doi.org/10.1080/00207543.2015.1037934

Wang, L., Ma, C., Feng, X., et al. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, *18*, 186345. https://doi.org/10.1007/s11704-024-40231-1

White, S. K. (2025, May 20). *What is SCOR? a model to improve supply chain management*. CIO. https://www.cio.com/article/222381/what-is-scor-a-model-for-improving-supply-chain-management.html

Wieland, A., & Durach, C. F. (2021). Two perspectives on supply chain resilience. *Journal of Business Logistics*, *42*(3), 315–322. https://doi.org/10.1111/jbl.12271

Wijnands, S., Sharpanskykh, A., & Aly, K. (2024). *Generation of synthetic aircraft landing trajectories using generative adversarial networks* [Final published version; published in 14th SESAR Innovation Days (SIDS 2024)]. https://resolver.tudelft.nl/uuid:e7b09baf-7c44-471a-85bf-d0e668443e28

Wilkinson, S. (1998). Focus group methodology: A review. *International Journal of Social Research Methodology*, *1*(3), 181–203. https://doi.org/10.1080/13645579.1998.10846874

Wilson, F. (2013). The discipline of creativity [Uploaded as `the-discipline-of-creativity.pdf`].

Wu, T., Li, J., Bao, J., & Liu, Q. (2024). Processcarbonagent: A large language models-empowered autonomous agent for decision-making in manufacturing carbon emission management. *Journal of Manufacturing Systems*, *76*, 429–442. https://doi.org/10.1016/j.jmsy.2024.08.008

Xiao, H., Ai, X., Song, G., Sun, Y., & Shi, C. (2024). A comprehensive survey of large language models and multimodal llms. *Information Fusion*, *103*, 102328. https://doi.org/10.1016/j.inffus.2024.102328

Xu, A., Yu, T., Du, M., Gundecha, P., Guo, Y., Zhu, X., Wang, M., Li, P., & Chen, X. (2024). Generative ai and retrieval-augmented generation (rag) systems for enterprise. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 5599–5602.

Xu, Z., Cruz, M. J., Guevara, M., Wang, T., Deshpande, M., Wang, X., & Li, Z. (2024). Retrieval-augmented generation with knowledge graphs for customer service question answering. *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, 2905–2909.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd). Sage.

Yu, J., Zhu, H., & Zhao, X. (2025). Modular ai agents for transportation surveys and interviews. *Travel Behaviour and Society*, *40*, 100057. https://doi.org/10.1016/j.tbs.2025.100057

Zamani, E. D., Smyth, C., Gupta, S., & Dennehy, D. (2023). Artificial intelligence and big data analytics for supply chain resilience: A systematic literature review. *Annals of Operations Research*, *327*(2), 605–632. https://doi.org/10.1007/s10479-022-04983-y

Zhang, Q., Li, Z., Zhou, X., et al. (2025). The memory mechanism of large language models: A survey. *ACM Computing Surveys*. https://doi.org/10.1145/3748302

Zhang, W., & Zhang, J. (2025). Hallucination mitigation for retrieval-augmented large language models: A review. *Mathematics*, *13*(5), 856.

Zhao, S., Hong, J., & Lau, R. S. M. (2023). Impact of supply chain digitalization on supply chain resilience and performance. *International Journal of Production Economics*, *257*. https://doi.org/ https://doi.org/10.1016/j.ijpe.2023.108817

Zhong, T., Liu, Z., Pan, Y., Zhang, Y., Zhou, Y., Liang, S., Wu, Z., Lyu, Y., Shu, P., Yu, X., et al. (2024). Evaluation of openai o1: Opportunities and challenges of agi. *arXiv preprint arXiv:2409.18486*.

Zhou, H., Benton, W. C., Schilling, D. A., & Milligan, G. (2011). Supply chain integration and the SCOR model. *Journal of Business Logistics*, *32*(4), 332–344. https://doi.org/https://doi.org/10.1111/j.0000-0000.2011.01029.x

A

First Interview Iteration

A.1. Interview 1 - AI Expert

Interview Summary

Based on the interview, the expert frames Artificial Intelligence (AI) as the broad umbrella, with Generative AI (GenAI) and Agentic AI as key subsets. The expert distinguishes Agentic AI by its unique capability to formulate its own plan, act, and validate its own outcomes.

Despite this autonomy, the expert stresses that human-in-the-loop (HITL) oversight is essential today to manage outputs and align with corporate principles. This human element is the primary safeguard against key failure modes like hallucination and incorrect input data, which can occur even in agentic systems. Other critical safeguards include robust data protection (vital for enterprises) and building safety in by design, using root-cause analysis to manage errors when they happen.

Regarding implementation, the expert notes that while organizations set autonomy levels, high risk-aversion is currently slowing adoption. Before scaling autonomy, two prerequisites are critical: mature data management (to avoid ”garbage in, garbage out”) and employee AI maturity—retraining staff to move beyond low-value tasks toward creative, high-value prompting.

For pilot projects, the expert advises focusing on end-to-end business processes rather than single tasks, targeting repetitive work requiring extensive orchestration for the highest value. This approach is operationally costlier than simpler GenAI copilots, as agentic solutions require more models and complex orchestration to solve these end-to-end problems. However, the value is in achieving greater efficiency and quality, ultimately freeing human workers for higher-impact roles. The expert also confirmed that Agentic AI excels at complex data reconciliation and proactive disruption management in the supply chain context.

Open Coding

Table A.1: Open Coding Analysis – First Interview (AI Expert)

Code Name	Definition	Source Extract
Artificial Intelligence (AI)	The broad concept encompassing machine learning and models used to mimic human behavior.	”AI is the broader umbrella of concepts in which machine learning and various models are used to mimic aspects of human behavior.”
Machine Learning	A component within the broader AI umbrella used alongside various models to mimic aspects of human behavior.	”AI is the broader umbrella of concepts in which machine learning and various models are used to mimic aspects of human behavior. Within that umbrella are machine learning, Generative AI (GenAI), and Agentic AI, which are considered subfunctions or subsets of AI concepts.”
Subsets of AI	Generative AI (GenAI) and Agentic AI, which exist under the broad AI umbrella.	”Within that umbrella are machine learning, Generative AI (GenAI), and Agentic AI, which are considered subfunctions or subsets of AI concepts.”
Agentic AI (Definition)	Distinguished by its ability to independently formulate a plan, act, and verify the results by shaping its own approach and validating outcomes.	”Agentic AI differs in that it can shape its own approach and validate its own outcomes, meaning it can formulate a plan, act, and then verify the results.”
Human Oversight (HITL)	Necessary for every task today, requiring a human-in-the-loop (HITL) to monitor or fully validate outputs to ensure quality aligns with corporate principles.	”The human perspective is essential for every task, whether run by an Agentic AI or a GenAI application, and human oversight is crucial today. In almost all cases, a robust monitoring system is needed, implemented through alerts or through full validation of the outputs. This requires a human-in-the-loop to ensure that output quality meets expectations and aligns with corporate principles.”

Continued on next page.

Table A.1 continued.

Code Name	Definition	Source Extract
Failure Mode: Hallucination	A key failure mode that can occur even in Agentic AI, leading to incorrect overall results, including fabricated numbers, even if the agent attempts to validate outcomes.	”Key failure modes include hallucination and the use of incorrect input data. Hallucination can also occur in Agentic AI: even if an agent shapes its approach and validates outcomes, the process can still produce an incorrect overall result, including fabricated numbers.”
Failure Mode: Input Quality	Poorly framed inputs or queries that increase the chance of unsatisfactory results. This can lead to the generation of spurious figures.	”When a system is allowed to perform its own calculations, it may generate spurious figures, and poorly framed inputs or queries can further increase the chance of unsatisfactory results, with many things in the main process potentially leading to model failure.”
Safeguard: Data Protection	Critical for large organizations to prevent data misuse due to confidentiality obligations, regulatory requirements, and customer relationships, ensuring systems adhere to guidelines.	”Data protection is a critical safeguard. Large organizations must prevent data from reaching the wrong hands or unintended model providers because of confidentiality obligations, customer relationships, and regulatory requirements. Systems should be built to adhere to government and company guidelines so data is protected and used appropriately.”
Safeguard: Human Foundation	Maintaining system integrity by keeping a human in the loop and embedding monitoring to continuously observe and adjust behavior over time.	”Another key safeguard is maintaining a human foundation for generated information, which means keeping a human in the loop and embedding good monitoring so organizations can continuously observe system behavior and adjust over time.”
Setting Autonomy Levels	Determined by governments (e.g., EU bodies) or corporations via AI strategies, with policies defined per use case concerning models and data protection.	”Autonomy levels are generally set by organizations and governments—such as EU bodies—or by corporations through their AI strategies. Policies are defined per use case, specifying the principles that apply to models and data protection, and policy needs to be assessed level by level.”

Continued on next page.

Table A.1 continued.

Code Name	Definition	Source Extract
Impact of Risk Aversion	High corporate risk aversion limits small experiments, proofs of concept, and demos, which subsequently slows the adoption curve.	”A current complication is that many corporations are highly risk-averse and maintain strict policies. While this is often positive, it also limits experimentation, small proofs of concept, and demos, which slows the adoption curve across projects.”
Exception Management Principle	Safety and exception-management principles must be built in by design, engineering systems to prevent exceptions.	”Safety and exception-management principles should be built in by design. [...] Systems should be engineered so that exceptions are prevented, which requires knowledge of how to configure AI systems, awareness of common pitfalls, and development practices that avoid undesirable modes.”
Undesirable Modes	The negative system behaviors that skilled development and proper configuration of AI systems must aim to avoid through specific development practices.	”Systems should be engineered so that exceptions are prevented, which requires knowledge of how to configure AI systems, awareness of common pitfalls, and development practices that avoid undesirable modes.”
Skilled Development Requirement	Essential for preventing exceptions, ensuring developers can instruct the AI effectively and specify the correct features, especially when GenAI assists in application creation.	”Skilled development is essential, and builders must be able to instruct the AI effectively and specify the right features when GenAI assists in creating the application.”
Error Recovery Strategy	The most important strategy is root-cause analysis, tracing system behavior across steps to assess correctness, improve the system, and enhance explainability, while maintaining the human in the loop.	”It is crucial to maintain the human in the loop and preserve human understanding. The most important strategy is root-cause analysis: by tracing how the system behaved across its steps, stakeholders can assess the correctness of process stages and outputs, learn how to improve the system, and enhance explainability.”

Continued on next page.

Table A.1 continued.

Code Name	Definition	Source Extract
Value KPI: Efficiency & Quality	Agentic AI is implemented to achieve greater efficiency (faster than a human) or better quality (broader scope/-greater scalability), aiming to allow humans to focus on core strengths.	”Agentic AI is implemented to do things better or more efficiently. ‘Better’ refers to a broader scope of tasks and greater scalability, while ‘efficient’ refers to completing work faster than a human. The goal is to free humans to focus on what they do best.”
Value KPI: Employee Utilization/Satisfaction	Key performance indicators include employees accomplishing more within roles (utilization) and being happier at work (satisfaction), which leads to delivering more impactful results.	”Key indicators include employee utilization—people should be able to accomplish more within their roles—and satisfaction, as people should be happier at work and deliver more impactful results. Value also depends on fostering creativity rather than just technical skill, since humans still provide creative input through prompting and instruction.”
Cost Driver: GenAI Copilots	Generally cheaper operationally, assisting with quick help, simple queries, and smaller tasks, often invoking only one large language model.	”GenAI copilots typically assist with smaller tasks, quick help, and simple queries, and they are often cheaper because they may invoke only one large language model.”
Cost Driver: Agentic AI (Operational)	Costlier operationally because it handles complex, end-to-end problems, often requiring more prompts and sometimes multiple models to reach a decision. This relates to the differentiation of Unit Economics.	”Agentic AI is applied to more complex, end-to-end problems and often produces more creative outputs because it shapes its approach and validates outcomes; operationally it tends to be costlier, as it usually requires more prompts and sometimes multiple models to reach a decision.”
Cost Driver: Agentic AI (Implementation)	Higher implementation costs due to the need to define agent behaviors, choose the right tools, and add orchestration and monitoring.	”Implementation costs are also higher for Agentic AI because teams must choose the right tools, define agent behaviors, and add orchestration and monitoring.”

Continued on next page.

Table A.1 continued.

Code Name	Definition	Source Extract
Prerequisite: Data Management/Maturity	Organizations must know data sources and meanings; poor data (”garbage in, garbage out”) undermines outputs, and addressing this is costly, often explaining AI’s failure to meet expected value.	”First, data management and maturity are frequently lacking: organizations must know data sources and meanings, because ‘garbage in, garbage out’ applies, and poor inputs will undermine outputs; addressing data management is costly and often explains why AI falls short of expected value.”
Prerequisite: Employee AI Maturity	Employees currently use GenAI for low-value tasks and need training to become more creative in prompting and selecting use cases that deliver real value.	”Second, employee AI maturity remains limited: many people use GenAI for low-value tasks, whereas they need to be far more creative in prompting and in selecting use cases that deliver real value.”
Pilot Selection Criterion (End-to-End)	Pilots should focus on the entire business process, emphasizing linking tasks so downstream work benefits, rather than fixing only a single high-value step.	”Selection should focus on the end-to-end business process. Rather than fixing only the highest-value single step, the strategy emphasizes linking tasks and connecting use cases so downstream work benefits.”
Highest Value Use Cases	Found in repetitive work that requires extensive orchestration, where the goal is to replicate the entire process within GenAI solutions to achieve organizational flow change.	”The preferred approach is to replicate the process within GenAI solutions so the organizational flow changes as a whole, not just one isolated use case. Teams map processes, detail the steps, and identify where value is highest, which often occurs in repetitive work that requires extensive orchestration.”

Continued on next page.

Table A.1 continued.

Code Name	Definition	Source Extract
SC Benefit: Data Reconciliation	A prime application where Agentic AI excels in filling in missing data, harmonizing formats, validating correctness, and using authoritative systems to update information automatically.	”Greatest value appears where tasks are repetitive and require extensive orchestration, with data reconciliation as a prime example. [...] Agentic AI can fill in missing data, harmonize formats, and, when appropriate, use authoritative systems to update information automatically. This includes checking inputs along the process, validating correctness, and looping back to data providers when information is missing or wrong.”
SC Benefit: Proactive Disruption Management	Agentic AI serves as an orchestration mechanism across diverse data sources, monitoring for early signals, accessing core systems for investigation, and planning follow-ups to automate tracking and sense-making for major disruptions.	”Agentic AI serves as an orchestration mechanism across diverse data sources. For proactive disruption management, it can add significant value because potential disruptions arise from many places and are hard for humans to track continuously. An agent can monitor for early signals, access core systems to determine the next investigative steps, and use available tools to plan follow-ups. By harmonizing inputs across systems, it can automate tracking and sense-making for major disruptions and support faster, more adaptive responses.”

Themes and Supporting Open Coding Names

Table A.2: Themes and Supporting Open Coding Names – First Interview (AI Expert)

Theme Name	Supporting Open Coding Name
Theme 1: The AI hierarchy and agentic mechanism	Artificial Intelligence (AI), Subsets of AI, Agentic AI (Definition), Machine Learning
Theme 2: Governance, oversight, and safety-by-design	Human Oversight (HITL), Setting Autonomy Levels, Safeguard: Data Protection, Exception Management Principle, Error Recovery Strategy, Safeguard: Human Foundation, Undesirable Modes, Skilled Development Requirement

Continued on next page.

Table A.2 continued.

Theme Name	Supporting Open Coding Name
Theme 3: System failure modes and error generation	Failure Mode: Hallucination, Failure Mode: Input Quality
Theme 4: Value drivers and operational economics	Value KPI: Efficiency & Quality, Value KPI: Employee Utilization/Satisfaction, Cost Driver: GenAI Copilots, Cost Driver: Agentic AI (Operational), Cost Driver: Agentic AI (Implementation)
Theme 5: Organizational readiness and strategic pilot selection	Prerequisite: Data Management/Maturity, Prerequisite: Employee AI Maturity, Pilot Selection Criterion (End-to-End), Highest Value Use Cases, Impact of Risk Aversion
Theme 6: Agentic value in supply chain orchestration	SC Benefit: Data Reconciliation, SC Benefit: Proactive Disruption Management

A.2. Interview 2 - AI Expert

Interview Summary

AI is the broad field of systems that learn or reason; Generative AI (GenAI) learns data patterns to create new content; Agentic AI builds on GenAI to plan, act, and check results across multi-step tasks using external tools with some autonomy. Agentic workflows usually cost more to run because they chain models and tool calls; costs depend mainly on training versus inference, model size, data volume, and the depth of the workflow. Key risks remain - and often grow - in agentic setups: hallucination, bias from training data, data or memory poisoning, and input manipulation (prompt injection). Extra risks arise from coordination: agents can disagree, a single compromised component can disrupt the whole process, and handoffs between humans and AI can fail. Autonomy should match task risk: high-risk tasks keep humans in the loop and require systematic testing before and after deployment; routine low-risk tasks can allow more automation. During operation, exceptions should trigger reduced autonomy, escalation to humans, or compensating actions; tool failures are easier to spot, while plan drift needs explicit monitoring. Resilience needs detailed state logs for rollback, protected and backed-up memories and data, and clear recovery playbooks. Evaluation must cover both goal achievement (e.g., efficiency, accuracy) and risk reduction (accountability, fairness, robustness, transparency), with adversarial tests and monitoring for distribution shift. Organizational readiness requires a technical and safety function that turns ethical standards into testable checks, clear legal responsibility, and strong governance over data and access. Early agentic pilots should target low-risk, well-specified tasks with a reference process, validate performance and safeguards, then scale autonomy carefully.

Open coding

Table A.3: Open Coding Analysis – Second Interview (AI Expert)

Code Name	Definition	Source Extract
AI/GenAI Definition	Explains AI as the overall field and GenAI as a subfield focused on generating new content.	”AI (Artificial Intelligence) is the overarching term... Generative AI (GenAI) is one subfield within this broader field.”
GenAI Model Singularity	Notes that GenAI, in common discussions, generally refers to a single model type.	”GenAI refers to a single type of model.”
GenAI Function	Describes how GenAI models operate by learning the distribution of the data.	”learn the distribution of the data rather than decision boundaries.”
LLM Parameter Size	Contextual scale of Large Language Models (LLMs) used in GenAI.	”LLMs with billions or trillions of parameters.”
Agentic AI Definition & Structure	Defines Agentic AI as a GenAI-based system design coordinating multiple autonomous agents (frequently LLM-based) using external tools.	”autonomy to use external tools” / ”frequently LLM-based.”
Agentic AI Capability/Metaphor	Agentic AI handles multi-step tasks and is described as ”an army of AI models chained together.”	”intuitively understood as an ‘army of AI models chained together’.”
Agentic Use Case Example (Traffic)	The concrete scenario used to illustrate a potential Agentic AI application.	”traffic-light control as a concrete case.”
Traffic Control Agent Workflow	The multi-agent sequence used in the traffic control example, including roles like perception and optimization.	”extract demand from footage and then call another agent, an optimizer... then issue commands back to the control agent.”
GenAI Failure Mode: Hallucination	Uncontrolled compositional logic in GenAI leads to outputs violating knowledge/rules.	”Hallucination is frequent in large generative models.”
General AI Risk: Data Integrity	Malicious entries corrupting training or memory stores.	”Data pollution or memory poisoning remains a fundamental risk.”

Continued on next page.

Table A.3 continued.

Code Name	Definition	Source Extract
Input Manipulation/Adversarial Attack	Exploiting hidden cues (e.g., invisible text commands) so systems follow instructions humans miss.	”Input manipulation can exploit hidden cues.”
Agentic Specific Failures	Failures due to coordination and autonomy: contradiction, single-agent compromise, and agents fabricating answers to bypass human oversight.	”agents can contradict one another, adversaries can compromise a single agent” / ”agents fabricating answers to bypass oversight.”
Failure Detection Difficulty	Tool failure is easy to flag, but plan drift in multi-step plans is difficult to detect.	”Tool failure is usually straightforward to flag... Plan drift is harder to detect.”
Tool Failure Mechanism	The straightforward detection method for tool failure within an Agentic workflow.	”Tool failure is usually straightforward to flag because it appears as a missing link in the workflow.”
Non-Transparency Justification	Caution is needed because complex models are non-transparent, making surprises difficult to anticipate.	”complex models are not transparent... anticipate all ‘surprises’.”
Oversight Principle (Risk-Based)	Oversight level depends on task risk, not the AI type (GenAI vs. Agentic).	”Human oversight should be determined by the task and its risk.”
Autonomy/Risk Scaling	The level of autonomy must be inversely proportional to the task’s risk.	”Autonomy should scale with risk.”
High-Risk Scenario Oversight	Safety-critical tasks (like traffic control) require strong human oversight due to potential severe consequences.	”Safety-critical tasks with the possibility of severe consequences require strong human oversight.”
Risk Contextuality	System risk, and thus oversight, is scenario-dependent; even simple models can become high-risk if integrated into a critical Agentic workflow.	”level of oversight is context- and scenario-dependent.”
Oversight Adjustment Rule	A practical rule for setting oversight by comparing machine output to human competence and adjusting based on potential consequences.	”compare machine performance with competent human performance and adjust oversight according to potential consequences.”

Continued on next page.

Table A.3 continued.

Code Name	Definition	Source Extract
Low-Risk Autonomy Example	A concrete example of a task suitable for high autonomy (basic image recognition).	”low-risk tasks that are simple or already show better-than-human performance... can operate with higher autonomy.”
Complexity/Risk Scaling (Network)	Expanding the scope of a task (e.g., traffic control to network level) rapidly increases risk, demanding human control.	”network level, human control should remain in place as complexity and risks increase rapidly.”
Minimum Safeguard: Evaluation	Systematic pre-deployment evaluation and iterative testing to reveal errors/biases.	”Before deployment, models need systematic evaluation of outputs and iterative testing.”
Minimum Safeguard: Continuous Oversight	Safety-critical applications require continuous human oversight and human-in-the-loop checks post-deployment.	”continuous human oversight across the lifecycle... human-in-the-loop checks.”
Internal Mitigation Strategy	Using a separate agent (evaluator/critic) to mitigate risk when objective rules are difficult to specify.	”separate agent as an evaluator or critic.”
Cross-Validation Mechanisms	Additional safeguards for validation using internal agent critique or external database checks.	”cross-agent critique or from database verification.”
Bias Mitigation Method	Specific method for handling persistent bias derived from training data.	”mitigation often relies on post-processing rules or evaluations for sensitive content.”
Exception Management (Combined Strategy)	Managing exceptions requires a combination of degrading autonomy, escalating to humans, and compensating strategies.	”Degrading autonomy, escalating to humans, and compensating with alternative strategies should be combined.”
Error Recovery Mechanisms	Mechanisms like robust state logging (for rollback), strong data protection/backups, and compensating actions are needed for recovery.	”robust state logging to enable rollback... and compensating actions that trigger human intervention.”
Incident Learning	The role of ongoing learning and extensive testing in managing unpredicted operational issues.	”Extensive pre-deployment testing and ongoing learning from incidents help manage surprises.”

Continued on next page.

Table A.3 continued.

Code Name	Definition	Source Extract
KPI Structure (Two Dimensions)	KPIs must cover both objective achievement (performance) and risk minimization (robustness, fairness, transparency).	”The first is performance and objective achievement... The second is risk minimization... including... robustness.”
Robustness Checks	Robustness checks test resistance to malicious activity and monitor for distribution shift.	”robustness checks should probe resistance to malicious activity... and they should monitor for distribution shift.”
Agentic Cost: Computation	Agentic AI requires significantly more computation due to the coordination of multiple models/tools.	”Agentic AI typically consumes far more computation.”
Key Cost Drivers	Cost variability depends on training method (from-scratch vs. fine-tuning), base model parameters, and data volume.	”training method itself (from-scratch training is far more expensive than fine-tuning).”
Organizational Prerequisite: Safety Function	Requirement for a technical and safety function to translate standards into engineering checks.	”Organizations need a technical and safety function.”
Organizational Prerequisite: Governance	Requirements for clear legal frameworks (assigning responsibility) and governance/access controls (authentication/permissioning).	”clear legal frameworks that assign responsibility for harm... governance and access controls.”
Agentic Pilot Criteria	Pilots should target low-risk tasks with defined steps and existing reference processes to validate performance before granting full control, aiding safe scaling.	”low-risk tasks with clearly defined steps and an existing reference process... supports safe scaling.”

Themes and Supporting Open Coding Names

Table A.4: Themes and Supporting Open Coding Names – Second Interview (AI Expert)

Theme Name	Supporting Open Coding Name
Theme 1: Definitions, structure, and technical foundations of AI systems	AI/GenAI Definition; GenAI Model Singularity; GenAI Function; LLM Parameter Size; Agentic AI Definition & Structure; Agentic AI Capability/Metaphor; Agentic Use Case Example (Traffic); Traffic Control Agent Workflow
Theme 2: Risk-based oversight and autonomy	Oversight Principle (Risk-Based); Autonomy/Risk Scaling; High-Risk Scenario Oversight; Risk Contextuality; Oversight Adjustment Rule; Low-Risk Autonomy Example; Complexity/Risk Scaling (Network)
Theme 3: Failure modes and core technical risks	GenAI Failure Mode: Hallucination; General AI Risk: Data Integrity; Input Manipulation/Adversarial Attack; Agentic Specific Failures; Failure Detection Difficulty; Tool Failure Mechanism; Non-Transparency Justification; Bias Mitigation Method; Incident Learning
Theme 4: Safeguards, mitigation, and error recovery	Minimum Safeguard: Evaluation; Minimum Safeguard: Continuous Oversight; Internal Mitigation Strategy; Cross-Validation Mechanisms; Exception Management (Combined Strategy); Error Recovery Mechanisms
Theme 5: Organizational prerequisites	KPI Structure (Two Dimensions); Robustness Checks; Agentic Cost: Computation; Key Cost Drivers; Organizational Prerequisite: Safety Function; Organizational Prerequisite: Governance; Agentic Pilot Criteria

A.3. Interview 3 - AI Expert

Interview Summary

This summary clarifies that AI is the broad field of machine-learning for prediction or classification, GenAI is a narrower subset built on large language models that generate text, and Agentic AI layers an LLM with tools and actions to execute multi-step workflows with some autonomy. Use GenAI for judgment-heavy or high-stakes tasks with a human making the final call, and use Agentic AI for frequent, low-creativity work over large volumes of unstructured data where analysis must trigger actions. Autonomy should track risk: higher risk means tighter controls and more human oversight. Typical failures include hallucinations, weak math, limited memory, prompt injection, data exposure, and in agents the possibility that one wrong step cascades into many. Safe deployment relies on deterministic checks, human-in-the-loop reviews, and strong security for credentials and model access, plus clear stops, escalations, and rollbacks when checks fail. Costs are driven mainly by token usage and rise further with orchestration and API calls, so measure cost per use and latency against required accuracy. In supply chains, agents are most useful for reading and acting on document-heavy processes and for continuously scanning external signals to surface disruptions sooner and recommend timely adjustments.

Open coding

Table A.5: Open Coding Analysis – Third Interview (AI Expert)

Code Name	Definition	Source Extract
AI Definition	A broad set of machine-learning techniques used for prediction, classification, or modeling, encompassing simple methods like linear regression.	”AI is a broad set of machine-learning techniques used to predict, classify, or model for specific purposes; think of something as simple as linear regression.”
GenAI Definition	A narrower subset, typically Large Language Models (LLMs) within NLP, where words are encoded as numbers and modeled (often with neural networks) to generate responses.	”GenAI is a narrower subset—typically LLMs within NLP—where words are encoded as numbers and modeled (often with neural networks) to generate responses.”
Agentic AI Definition	A system using an LLM to take actions, reason over multiple steps, choose tools, interact with data/sensors, and execute non-linear workflows with autonomy.	”Agentic AI is different in that it uses an LLM to take actions: it accepts a user input, reasons over multiple steps, chooses tools, calls APIs, reads databases or local files, can interact with sensors, and executes non-linear workflows with some degree of autonomy.”
GenAI Suitable Tasks (Subjective Judgment)	Tasks involving subjective judgment or sensitive, high-stakes determinations requiring a human decision at the end (e.g., financial-crime risk assessments).	”Tasks that involve subjective judgment or sensitive, high-stakes determinations should stay under GenAI with a human in the loop; for example, financial-crime risk assessments may be data-driven but still require a human decision at the end.”
Agentic AI Suitable Tasks (Repetitive/Data-Driven)	Repetitive, low-creativity work in contexts dominated by large volumes of text or other unstructured data where the agent can read, analyze, and act.	”Agentic AI fits repetitive, low-creativity work that occurs often, and decision contexts dominated by large volumes of text or other unstructured data where the agent can read, analyze, and act.”
Oversight Principle (Risk/Subjectivity)	Oversight must increase as task subjectivity and risk rise, requiring more human review or deterministic safeguards for consequential steps.	”Oversight should increase as subjectivity and risk rise; the more technical or consequential the step, the more human review or deterministic safeguards are needed.”

Continued on next page.

Table A.5 continued.

Code Name	Definition	Source Extract
LLM Failure Mode: Hallucination	LLMs are probabilistic and designed to always answer, even when unsure.	”LLMs hallucinate because they are probabilistic and are designed to always answer, even when unsure.”
LLM Failure Mode: Cognitive Limits	Lack of true critical reasoning, limited long-term memory in long sessions, knowledge cut-off constraints, and weakness in exact mathematics.	”they lack true critical reasoning, have limited long-term memory in long sessions, and are constrained by a knowledge cut-off. They are also weak at exact mathematics.”
Operational Risk: Cloud Call Exposure	The specific context during which sensitive data may be inadvertently exposed.	”sensitive data may be exposed during cloud calls or output handling.”
Operational Risk: Prompt Injection	System instructions can be overridden by a security vulnerability.	”Operationally, prompt injection can override system instructions.”
Agentic Failure Mode: Extreme Agency	Amplification of errors where an LLM-driven misstep cascades into wrong subsequent actions (e.g., incorrectly changing production levels).	”At the agent level, ‘extreme agency’ can amplify errors—an LLM-driven misstep can cascade into wrong actions (e.g., incorrectly changing production levels).”
Safeguard: Model Limit Awareness	Start with awareness of model limits, respecting knowledge cut-offs, chunking long inputs, and verifying technical outputs.	”Start with awareness of model limits: respect knowledge cut-offs, chunk long inputs, and verify technical outputs.”
Safeguard: Deterministic Steps & Human-in-the-Loop	Inserting deterministic steps for exact calculations and placing a human in the loop where judgment is necessary.	”Insert deterministic steps for exact calculations and place a human in the loop where judgment is needed.”
Safeguard: Security Posture & Credentials	Protecting data via strong cybersecurity, vetted and protected API keys, approved model providers, and operating within an enterprise secure framework.	”Protect data through strong cybersecurity, vetted and protected API keys, and approved model providers.” / ”In enterprise settings, operate within a pre-approved, secure framework for models and credentials.”
Safeguard: Agent Controls	Defending against prompt injection and adding controls around the agent’s tool use and outputs.	”Defend against prompt injection and add controls around tool use and outputs.”

Continued on next page.

Table A.5 continued.

Code Name	Definition	Source Extract
Autonomy/Risk Matching Principle	Autonomy must be matched to risk and determinism: the higher the risk or subjectivity, the lower the autonomy and the stronger the controls.	”Match autonomy to risk and determinism: the higher the risk or subjectivity, the lower the autonomy and the stronger the controls.”
Control Improves Precision	Tighter control makes the agent’s behavior more reliable.	”In practice, tighter control improves precision—the more controlled the agent, the more reliable its behavior.”
Exception Goal: System Stability	The outcome of effective exception management, which includes graceful degradation, timely escalation, and safe compensation.	”Together, they enable graceful degradation,” / ”timely escalation, and safe compensation.”
Exception Management: Agency over Agency	A second agent reviewing each LLM-involved decision to approve or reject it, catching anomalies and drift before they propagate.	”‘Agency over agency’ has a second agent review each LLM-involved decision to approve or reject it, catching anomalies and drift before they propagate.”
Exception Management: Deterministic Control	Coded guards (range checks, schema checks, or invariant tests) that accept valid outputs or route out-of-bounds results to a human/halt tools.	”Deterministic control adds coded guards—range checks, schema checks, or invariant tests—that accept valid outputs, route out-of-bounds results to a human, or halt and retry tools.”
Recovery Strategy: Hard Stops and Handoffs	Embedding control mechanisms to pause the pipeline and trigger a human-in-the-loop review when a deterministic check fails, preventing further actions.	”Embed hard stops and handoffs at key points in the workflow. When a deterministic check fails, pause the pipeline, trigger a human-in-the-loop review, and prevent further actions.”
Recovery Strategy: Explicit Thresholds	Using defined thresholds to halt or revert proposed changes (e.g., production adjustments) until re-validation occurs.	”Use explicit thresholds to halt or revert proposed changes (e.g., production adjustments) and require revalidation before resuming.”
Rollback Principle: Controls at Every Critical Step	The core philosophy for recovery strategies to ensure erroneous actions cannot propagate.	”The core idea is to place controls at every critical step so erroneous actions cannot propagate.”

Continued on next page.

Table A.5 continued.

Code Name	Definition	Source Extract
KPI: Cost per use (Tokens)	Primary Key Performance Indicator measured in input plus output tokens, balancing cost, speed, and accuracy.	”Cost per use measured in tokens (input plus output) and latency to result are primary, with a deliberate quality-versus-speed trade-off depending on the chosen model and task.”
KPI: Latency to Result	Primary Key Performance Indicator measuring the turnaround time to obtain a result.	”Cost per use measured in tokens (input plus output) and latency to result are primary, with a deliberate quality-versus-speed trade-off depending on the chosen model and task.”
KPI Trade-off: Quality vs Speed	The necessary consideration in measuring KPIs, where higher quality may mean longer runtimes and higher token consumption.	”with a deliberate quality-versus-speed trade-off depending on the chosen model and task.”
KPI Consideration: Accuracy Level Required	The specific requirement for the use case that must be weighed against cost and turnaround time.	”the KPI view should therefore weigh cost, turnaround time, and the required accuracy level for the use case.”
Cost Driver: Orchestration/Infrastructure	Costs incurred by agentic solutions beyond the LLM itself, including additional API calls, orchestration, and infrastructure.	”Agents also incur costs beyond the LLM itself—additional API calls, orchestration, and infrastructure—”
Cost Driver: Token Consumption Scaling	Drives variable cost and scales with depth of reasoning, retries, and tool use.	”Token consumption drives variable cost, and it scales with depth of reasoning, retries, and tool use.”
Cost Driver: Agency Over Agency Cost	Using a second agent review (”agency over agency”) roughly doubles the number of calls and therefore the cost.	”Agency over agency roughly doubles calls and therefore cost.”
Agentic Cost Outcome: Equal or Higher	Typical cost comparison of agentic solutions relative to simple GenAI prompts, while delivering more autonomy.	”so agentic solutions tend to be equal or higher in cost than simple GenAI prompts, while delivering more autonomy.”
Prerequisite: Risk Acceptance & Prevention	Teams must understand LLM and agent risks, accept potential failure, and put prevention first.	”Teams deploying agents must understand LLM and agent risks, accept that failure is possible, and put prevention first.”

Continued on next page.

Table A.5 continued.

Code Name	Definition	Source Extract
Prerequisite: Robust Cybersecurity Posture	An essential organizational requirement for safely deploying agents, including approved models, protected API keys, and a secured operational framework.	”A robust cybersecurity posture, approved models, protected API keys, and a secured operational framework are essential.”
Prerequisite: Governance for Access	Access to models should be governed through formal requests and provider agreements so that tooling and data flows are pre-approved.	”Access to models should be governed through formal requests and provider agreements so that tooling and data flows are pre-approved.”
Pilot Selection: Data Type	Prioritize domains with many unstructured documents or text-heavy decisions.	”Prioritize domains with many unstructured documents or text-heavy decisions.”
Pilot Selection: Task Nature	Tasks should be repetitive and low-creativity where automation meaningfully reduces effort.	”and with repetitive, low-creativity tasks where automation meaningfully reduces effort.”
Pilot Selection: Business Case	A clear business case must exist to justify costs, especially for always-on monitoring or research-heavy agents.	”Ensure there is a clear business case to justify costs, especially for always-on monitoring or research-heavy agents.”
SC Benefit: Document-Driven Action	Areas dependent on reading, summarizing, and acting on large bodies of text/documents, or requiring frequent, repetitive decisions.	”Areas that depend on reading, summarizing, and acting on large bodies of text or documents, or that require frequent, repetitive decisions, stand to gain the most.”
SC Benefit: Cadence Shift	Operational reviews can shift from periodic to daily cadence, converting unstructured data into timely actions.	”This includes document-driven assessments, ongoing operational reviews that can shift from periodic to daily cadence, and any stage where unstructured data must be turned into timely actions.”
SC Resilience: Continuous External Scanning	Agents continuously scan external signals (like news or maritime data) to detect emerging disruptions.	”Agents can continuously scan external signals—like news or maritime data—to detect emerging disruptions.”

Continued on next page.

Table A.5 continued.

Code Name	Definition	Source Extract
SC Resilience: Dynamic Monitoring and Response	Monitoring indicators (ship density, average speed) near chokepoints to recommend route or plan adjustments faster than periodic reviews.	”monitor indicators such as ship density or average speed near chokepoints, and recommend route or plan adjustments faster than periodic reviews.”
SC Requirement: Cost and Engineering Acknowledgment	The caveat regarding the effort and expense required to operate dynamic monitoring tools at scale.	”while acknowledging the cost and engineering required to operate such tools at scale.”

Themes and Supporting Open Coding Names

Table A.6: Themes and Supporting Open Coding Names – Third Interview (AI Expert)

Theme Name	Supporting Open Coding Name
Theme 1: AI taxonomy and core mechanisms	AI Definition; GenAI Definition; Agentic AI Definition
Theme 2: Operational risks and limits	LLM Failure Mode: Hallucination; LLM Failure Mode: Cognitive Limits; Operational Risk: Prompt Injection; Operational Risk: Cloud Call Exposure; Agentic Failure Mode: Extreme Agency
Theme 3: Control, safety, and oversight architecture	Autonomy/Risk Matching Principle; Oversight Principle (Risk/Subjectivity); Control Improves Precision; Safeguard: Model Limit Awareness; Safeguard: Deterministic Steps & Human-in-the-Loop; Safeguard: Security Posture & Credentials; Safeguard: Agent Controls; Prerequisite: Risk Acceptance & Prevention; Prerequisite: Robust Cybersecurity Posture; Prerequisite: Governance for Access
Theme 4: Exception management and recovery	Exception Management: Agency over Agency; Exception Management: Deterministic Control; Exception Goal: System Stability; Recovery Strategy: Hard Stops and Handoffs; Recovery Strategy: Explicit Thresholds; Rollback Principle: Controls at Every Critical Step
Theme 5: Economics, planning, and suitability	GenAI Suitable Tasks (Subjective Judgment); Agentic AI Suitable Tasks (Repetitive/Data-Driven); Pilot Selection: Data Type; Pilot Selection: Task Nature; Pilot Selection: Business Case; KPI: Cost per use (Tokens); KPI: Latency to Result; KPI Trade-off: Quality vs Speed; Cost Driver: Token Consumption Scaling; Cost Driver: Orchestration/Infrastructure; Cost Driver: Agency Over Agency Cost; Agentic Cost Outcome: Equal or Higher

Continued on next page.

Table A.6 continued.

Theme Name	Supporting Open Coding Name
Theme 6: Supply chain applications and resilience	SC Benefit: Document-Driven Action; SC Benefit: Cadence Shift; SC Resilience: Continuous External Scanning; SC Resilience: Dynamic Monitoring and Response; SC Requirement: Cost and Engineering Acknowledgment

A.4. Interview 1 - Supply Chain Expert

Interview Summary

The interview describes the Availability Leader as a highly demanding role at the end of the supply chain (Make, Deliver, Enable), combining proactive checks on stock availability for major promotions with reactive, last-minute mitigation when problems arise. The role sits between customers and planners in Poland, translating market issues into concrete planning actions. Daily work follows a fixed risk-assessment cycle before an 11:20 order cut-off, using a centralized European risk tool and escalating only substantial issues (such as missing around ten pallets). Typical disruptions include traffic delays, truck and rail failures, while major shocks (e.g., geopolitical events and transport bottlenecks) require significant production shifts and contingency measures. Resilience relies on quickly reallocating volume to other European plants and sourcing from other markets' stock and safety stock when label requirements allow.

Decision-making is guided by the KPI, which measures the share of assortment not replenished to safety stock; when it exceeds 5%, major service issues in the following week are considered almost certain. A key structural challenge is misaligned incentives and KPIs across functions, leading teams to optimise their own metrics rather than overall profitability and thereby distorting inputs for decisions.

AI is not yet used to directly manage core supply chain disruptions. Instead, an internal AI tool, *Inside*, mainly supports organizational learning, helping employees decode company-specific language and rehearse difficult conversations. The automation roadmap focuses on repetitive, rule-based tasks: codifying decision trees so that standard risk cases are handled automatically, and standardising excess-stock analysis via clear monetary thresholds. For more agentic behaviour, the system would need financial guardrails and approval workflows, with AI copilots proposing planning scenarios and AI agents executing routine reallocations and rescheduling within predefined limits.

Open coding

Table A.7: Open Coding Analysis – First Interview (Supply Chain Expert)

Code Name	Definition	Source Extract
Late-Stage Stock Mitigation	The primary function of the role, involving immediate, last-minute intervention to resolve stock problems in the market.	In practice, in the mornings I am usually busy with last-minute mitigation plans for stock. I work on everything that is going wrong in the market and try to solve it at the latest possible moment.
Intra-Cluster Shuttle Delay	A specific, common logistics disruption involving delayed transfers between internal company sites (e.g., from Germany to Belgium).	This can include, for example, a shuttle delayed from Germany to Belgium, where we distribute from.
Customer-Side Delay Negotiation	Operational response to logistics delays, focusing on arranging alternative deliveries or negotiating delivery time adjustments with the customer.	I check whether there is anything I can do on the customer side, such as delivering a few hours later, or arranging another priority shuttle from somewhere else, in order to have the product on time and avoid impact on customers.
Proactive Promotion Volume Check	Anticipatory work involving verifying production volumes are completed in time, triggered by large known customer promotions in the region.	When I see big promotions at certain customers in the France–Benelux region, I check whether we produced the required volumes in time.
Customer-Planner Bridging Role	The strategic position of the availability leader acting as a link between downstream customer demand and upstream production planning teams.	I call daily with planners in Poland, who manage capacity and production planning, and I position myself between customers and planners.
Non-Performing Stock Depletion	Responsibility for managing obsolete inventory from prior assortments, requiring alignment with the sales team to plan its removal.	I am responsible for old stock and non-performing stock from the old assortment, aligning with sales on depletion plans and ensuring that we do not have excessive stock in the distribution centres.
Geopolitical Conflict Production Loss	A major disruption triggered by war, necessitating immediate alternative production solutions to maintain service levels.	One major disruption was the war in Ukraine. We have a plant there, and we suddenly had to find another solution to keep producing stock on time.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
Cross-Product Capacity Doubling	The responsive strategy of diverting production volume to a plant that normally makes different products, forcing that plant to double its output.	We shifted volume to a plant in Germany, where other products are made, and that plant had to double its capacity to cover all the volume previously produced in Ukraine.
Common Truck Scheduling Failure	The routine operational issue where trucks fail to arrive at their planned time, requiring the incorporation of buffer time.	Traffic delays are very common. Trucks are often planned to arrive at a specific time and do not arrive as scheduled, so we must incorporate buffer time.
DC Congestion & Unloading Constraint	Issues arising at distribution centers when insufficient infrastructure capacity leads to trucks waiting and congestion building up.	Even when trucks are on time, the distribution centre may not have capacity to unload and reload, or the DC may already be full, resulting in trucks waiting and congestion building up.
Pilferage/Security Quality Return	A specific cause for inventory return where goods must be sent back because a transport vehicle was broken into, compromising product quality.	Another issue arises when goods return because a truck was broken into and the quality of the products can no longer be trusted.
Canal Blockage Promotion Loss	A significant logistical disruption involving major infrastructure failures (like a ship stuck in a canal) that leads to missed customer promotions.	Another significant disruption was the incident of a large ship stuck in a canal, which carried many of our products. That led to missed promotions and forced us to reconsider routing and supply alternatives.
Urgent Rail-to-Truck Shift	The mitigation necessity for critical items due to the high failure rate of scheduled rail transport, requiring immediate switching to road transport.	Rail shipments fail more frequently than they succeed, so we often need to consider urgent truck transport for the most critical items.
Cardboard Material Lead Time Bottleneck	A frequent supply chain constraint caused by extended lead times specifically for packaging materials.	This often leads to long lead times for packaging materials such as cardboard.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
Tier-N Supplier Dependency Risk	Vulnerability derived from reliance on extended international supply chains (Asia/Latin America) and dependence on suppliers multiple levels removed.	For some categories... the dependence on long international supply chains increases lead times. In those cases, we are dependent on the supply chain of our suppliers and even their suppliers.
Market Risk Input to Mitigation	The requirement for the availability leader to provide market context and prioritization to the Polish planning team's risk reports.	I provide the customer and market perspective on what is most important and how we should mitigate.
11:20 AM Risk Check Cut-off	The mandatory daily time constraint (before 11:20) for the concentrated activity of checking the risk dashboard.	This creates a limited time window in which I actively check all risks, which is before 11:20. In practice, my risk-checking activities are concentrated in the mornings.
Ten-Pallet Risk Filtering Threshold	The volume-based criterion (ten pallets) used by the planning team to escalate issues deemed significant enough for management attention.	We do not invest time on minor risks such as missing a single box. For cases like missing ten pallets, they contact me.
Centralized European Risk Tool	The software solution used across Europe to visualize all risks, centrally managed by the Poland team.	We use a centralized tool that shows all risks across Europe, managed from Poland. All products at risk are visible in this tool, and for each identified risk we define a mitigation plan.
Fixed 11:20 AM Order Cut-off	The mandatory deadline for customer order placement to be included in the daily allocation cycle.	Customers can place orders only until 11:20; after that, no new orders are processed for that day, and any new orders move into the next day's cycle.
Stock Allocation ("The Release" Phase)	The formalized stage immediately following the cut-off, dedicated to distributing available inventory, critical during shortages.	After the order cut-off, there is a phase called the release, in which we check how to allocate the available stock.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
Widespread Cardboard Shortage Event	A severe, industry-wide disruption of packaging material supply caused by external factors (COVID-19 surge in parcel deliveries).	During COVID-19, there was a major cardboard shortage because of the surge in parcel deliveries... This affected not only our company but many companies, and multiple brands across the market were impacted.
Sustainability Limit on Material Switching	Long-term sustainability goals acting as a constraint against rapid, reactive material substitutions (e.g., reverting to plastic).	Switching back to plastic packaging in a short time frame was not a viable option due to sustainability strategies and the move towards more cardboard, so the problem remained widespread and persistent.
Recurrent Local Strike Pre-Scheduling	The strategy of mitigating predictable local disruptions (like French strikes) by front-loading deliveries using additional trucks beforehand.	Strikes in France are a more local and recurrent phenomenon. They allow some room for preparation, such as scheduling additional trucks before the strike days, since no deliveries will occur during the strike.
Full DC Buffer Stock Restriction	The operational constraint where limited physical distribution center capacity prevents the proactive storage of buffer inventory.	If a DC is already running at full capacity, it may be impossible to receive additional stock in advance as a buffer.
Unilateral Promotion Change (2-for-5)	A severe, unexpected increase in demand caused by a customer changing the promotional mechanics without warning.	However, when the promotion started, the supermarket unilaterally changed the promotion to ”buy two, get five.” As a consequence, the stock allocated for the entire week was sold out on the first day.
3–4 Week Recovery Impossibility	The fixed internal constraint that production lead times (3–4 weeks for capacity/materials) are too long to recover stock for short, unexpected promotional peaks.	We plan capacity monthly, three months ahead, so if extra capacity, packaging or raw materials are needed suddenly, the required lead time of 3–4 weeks makes it impossible to recover stock for a one-week promotion.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
Cross-Country Stock Sourcing (Label Check)	Seeking inventory from other markets, conditional on label compliance with regulatory or customer standards.	We then aligned with higher-level stakeholders to see whether stock could be sourced from other countries, provided that the language on the labels did not cause issues.
Borrowing Neighboring Safety Stock	The mitigation strategy of temporarily utilizing the reserved safety stock of another market to maintain service in the affected region.	The second step was to verify whether other markets had stock they did not immediately need, and whether we could temporarily use their safety stock to protect our own service and profit.
Overstock Depletion Promotion Tie-in	A mitigation strategy aligning with sales to integrate slow-moving or excess products into the current promotion to achieve dual benefits (service and depletion).	A third option involved coordination with sales to include other available products in the promotion that we either had in sufficient quantity or even wanted to reduce.
Packaging Material Trade-Off Risk	A high-level decision involving sacrificing future supply stability (risking 3 weeks OOS) for immediate service recovery by using up limited materials.	This required evaluating a trade-off: depleting packaging materials to support the current promotion but risking three weeks of out-of-stock for other customers.
KPI: IWL Below (Safety Stock Coverage)	The primary metric used to gauge system robustness by tracking the percentage of assortment that has not achieved safety stock levels.	We use a KPI that directly reflects the risk of disruptions, called IWL below. It measures, as a percentage of our full assortment, how many items are replenished to their safety stock level.
5% IWL Critical Threshold	The established threshold (5%) for the IWL KPI that signals highly likely, significant service issues in the subsequent week.	If more than 5% of the assortment is not replenished to safety stock, significant issues in the following week are almost certain.
Scrapping Threshold (<$1,000)	The financial rule for low-value excess stock management: disposal is the most efficient action.	If the value of the stock is below 1,000 USD, scrapping it is often the most efficient option.
Broker Threshold (10k–50k)	The financial rule for medium-value excess stock: utilization of a third-party broker for disposition.	Between 10,000 USD and 50,000 USD, shipping to a broker can be more appropriate.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
Sales Alignment Threshold (>$50k)	The financial rule for high-value excess stock: mandatory collaboration with the sales team for commercial resolution.	Above 50,000 USD, alignment with sales to design a commercial action or a specific deal is usually required.
Decision Tree Automation Project	The current system development initiative to encode rules to allow autonomous processing of standard risk cases.	I am working on a detailed decision tree that will allow the system to handle most cases automatically.
Learning AI Evolution Roadmap	The long-term vision for system resilience: moving past static logic to AI that learns from history and applies solutions autonomously.	In the longer term, I see an AI-based evolution beyond a static decision tree, where the system learns from past situations and automatically applies similar solutions when comparable conditions arise.
European Multi-Plant Network Resilience	The structural advantage of maintaining a regional network of numerous manufacturing sites within Europe to enhance system resilience.	From a network design perspective, one of the strengths of a large FMCG manufacturer is having many plants located within Europe, despite operating globally. This structure supports resilience.
Local Footprint Impact Limitation	The benefit of the regional strategy: limiting the duration of severe production disruptions (like Ukraine) to a short period (e.g., two weeks).	When the plant in Ukraine stopped, we could shift production to Germany and continue with roughly two weeks of impact, rather than facing a multi-year disruption.
Must-Have: Promotion Visibility & Volume Shift	Enhanced promotional foresight and the operational flexibility to move stock internationally as key resilience factors.	A third important element is good forecasting and visibility for promotions, combined with the ability to shift volumes between markets.
Internal AI Tool (Confidentiality Rationale)	Development and use of an in-house AI solution to safeguard sensitive corporate data by avoiding public GenAI platforms.	We use an internal AI tool extensively because of the high level of confidential information we handle, which prevents us from using public tools like ChatGPT.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
AI Tool: Company Language Decoding	A core function of the internal AI tool: translating proprietary abbreviations and terminology for new employees and general use.	In meetings, especially for new employees, it helps decode the large number of internal abbreviations and company-specific terminology that form our ”company language.”
Organizational Blockage: Misaligned KPIs	Operational friction stemming from internal teams prioritizing their individual metrics over maximizing global company profit.	Blockages often arise not from technical data quality issues but from misaligned incentives and KPIs. Different teams try to optimize their own performance indicators rather than the overall company outcome.
KPI Protection Conservative Behavior	A specific instance of misaligned incentives: planners providing intentionally conservative raw material estimates to protect their metrics.	For instance, colleagues in Poland may state that raw materials cannot be brought in earlier, because they want to avoid the risk of missing them and negatively affecting their own KPIs.
AI Tool: Managerial Simulation/Training	Using the internal tool to practice sensitive managerial interactions (e.g., career conversations or negative feedback).	For example, when managing a direct report, I can use the AI tool to practice career conversations or difficult dialogues as a manager.
Automation Priority: Risk Call & Excess Stock	Two repetitive and rule-based processes identified as ideally suited for autonomous system management.	I see clear potential for automation in two areas: the daily risk call and the weekly excess stock analysis.
Decision Tree Bounded Execution	Potential for the automated system to execute standard mitigation steps, provided strict guardrails are in place, freeing human focus for exceptions.	With appropriate guardrails encoded in this logic, the system could autonomously propose or even execute standard mitigation actions, while I focus on the alerts and exceptional cases.
Automation Guardrail: Financial/-Workflow	Limits for autonomous action, defined by financial value thresholds and established human approval paths.	The guardrails would be financial thresholds and clear approval workflows for larger amounts.
Core Customer Collaboration Need	Fundamental requirement for improving supply chain resilience through partnership with customers: shared forecasting and order visibility.	With customers, the most important aspects are collaboration on forecasts and visibility on expected orders.

Continued on next page.

Table A.7 continued.

Code Name	Definition	Source Extract
Incentive: Full Truckload Rewards	A collaboration mechanism rewarding customers who order in efficient bulk quantities, enabling direct plant delivery and avoiding intermediate DC shuttling.	Incentive mechanisms can support this, such as rewards for always ordering full truckloads of a specific SKU, which may enable direct shipping from plants and avoid intermediate DC shuttling.
Incentive: Advance Order Guarantee	A mechanism where the customer places promotional orders early (e.g., one month), and the company guarantees fulfillment.	Another mechanism is agreeing that promotional orders are placed one month in advance in the system for certain campaigns, in exchange for a guarantee of full delivery.
Planning: High-Volume Data-Heavy Decisions	Planning work relies on high data volume and repetitive decisions, making it a strong automation target.	On the planning side, there is a very large volume of data, and many decisions are repetitive and data-heavy.
Copilot/Agent SC Planning Roles	Distinguishing future AI roles: copilots draft options/scenarios; agents autonomously execute standard reallocations and rescheduling within limits.	A copilot could be used to propose planning options and scenarios, while agents could execute standard reallocations, rescheduling and risk mitigations within predefined limits.

Themes and Supporting Open Coding Names

Table A.8: Themes and Supporting Open Coding Names - First Interview (Supply Chain Expert)

Theme Name	Supporting Open Coding Name
Theme 1: Role definition & structured workflow	Late-Stage Stock Mitigation; Customer-Planner Bridging Role; Market Risk Input to Mitigation; 11:20 AM Risk Check Cut-off; Ten-Pallet Risk Filtering Threshold; Centralized European Risk Tool; Fixed 11:20 AM Order Cut-off; Stock Allocation (”The Release” Phase); Non-Performing Stock Depletion
Theme 2: Logistics, external & demand shocks	Intra-Cluster Shuttle Delay; Common Truck Scheduling Failure; DC Congestion & Unloading Constraint; Pilferage/Security Quality Return; Canal Blockage Promotion Loss; Urgent Rail-to-Truck Shift; Widespread Cardboard Shortage Event; Full DC Buffer Stock Restriction; Unilateral Promotion Change (2-for-5)

Continued on next page.

Table A.8 continued.

Theme Name	Supporting Open Coding Name
Theme 3: Crisis mitigation and resource constraint	Customer-Side Delay Negotiation; Proactive Promotion Volume Check; Geopolitical Conflict Production Loss; Cross-Product Capacity Doubling; Cardboard Material Lead Time Bottleneck; Tier-N Supplier Dependency Risk; 3–4 Week Recovery Impossibility; Cross-Country Stock Sourcing (Label Check); Borrowing Neighboring Safety Stock; Overstock Depletion Promotion Tie-in; Packaging Material Trade-Off Risk
Theme 4: Resilience structure and inventory metrics	European Multi-Plant Network Resilience; Local Footprint Impact Limitation; Must-Have: Promotion Visibility & Volume Shift; KPI: IWL Below (Safety Stock Coverage); 5% IWL Critical Threshold; Scrapping Threshold (<1,000); Broker Threshold (10k–$50k); Sales Alignment Threshold (>$50k)
Theme 5: Organizational friction and external partnership	Organizational Blockage: Misaligned KPIs; KPI Protection Conservative Behavior; Core Customer Collaboration Need; Incentive: Full Truckload Rewards; Incentive: Advance Order Guarantee; Sustainability Limit on Material Switching; Recurrent Local Strike Pre-Scheduling
Theme 6: Automation, AI tools, and future agents	Decision Tree Automation Project; Learning AI Evolution Roadmap; Internal AI Tool (Confidentiality Rationale); AI Tool: Company Language Decoding; AI Tool: Managerial Simulation/Training; Automation Priority: Risk Call & Excess Stock; Decision Tree Bounded Execution; Automation Guardrail: Financial/Workflow; Planning: High-Volume Data-Heavy Decisions; Copilot/Agent SC Planning Roles

A.5. Interview 2 - Supply Chain Expert

Interview Summary

The interviewee is a manager in a sustainable supply chains team within a global consulting firm in the Netherlands, working across strategy, supply chain, and sustainability and effectively sitting on top of the SCOR processes. In the past few years, the main disruptions have been the pandemic, which halted global flows, drove transport prices up, and triggered reflection on nearshoring and dual sourcing, and new sustainability regulations, which forced companies to reassess supplier bases linked to issues such as deforestation; natural disasters also created shocks, especially at deep-tier suppliers, with effects propagating downstream. There is no single proactive system that consistently flags issues, although the interviewee has seen an advanced AI-based risk-analysis tool used by an engineering consultancy that maps risks across multiple tiers and provides a prioritized risk list for management, mainly in indirect procurement. Their own involvement is mostly conceptual, focused on governance and sourcing strategy, including upstream third-party risk management, while procurement decisions often remain dominated by cost considerations because the strategic value of resilience is hard to communicate to senior management. To improve resilience, the interviewee stresses visibility, transparency, and ultimately traceability in the upstream supply base, with traceability enabling active mitigation of value-

chain issues. Most tools for resilience target this upstream side, combining AI with external and trade data or geospatial and image-recognition techniques, but they face persistent challenges around data quality and integration, often addressed through probabilistic estimation. Data sharing along the chain is seen as the most powerful lever for resilience, yet many companies hesitate to share supplier data because they view it as a source of competitive advantage, and when thinking about agentic or copilot support across SCOR, the interviewee would clearly prioritize the Source function, where most disruptions originate.

Open coding

Table A.9: Open Coding Analysis – Second Interview (Supply Chain Expert)

Code Name	Definition	Source Extract
SCM Strategy Overlay	Interviewee is a manager in sustainable supply chains, covering strategy and sustainability across all SCOR processes, sitting ”on top” of them.	my work sits primarily in strategy, supply chain, and sustainability. It is quite all-encompassing across the SCOR elements: rather than focusing on a single area, my role is more generic and sits ”on top” of the different SCOR processes.
COVID Sourcing Shift	The COVID-19 disruption halted supply chains, prompting strategic consideration of nearshoring and dual-sourcing key suppliers.	One major disruption was COVID-19. Global supply chains effectively came to a halt... This triggered a lot of operational and strategic thinking around questions such as: should we nearshore some manufacturing, and should we dual-source key suppliers?
Regulatory Supplier Risk	New sustainability regulations force companies to reassess their supplier base due to compliance issues like deforestation or state ownership.	A second disruption relates to new sustainability regulations. Many European companies are now struggling with their supplier base, for example when suppliers are state-owned... or operate in regions linked to deforestation.
Deep-Tier Bullwhip	Natural disasters typically disrupt deep-tier suppliers (Tier-3/4), with impacts spreading downstream via a bullwhip effect that is often limitedly dampened.	these events tend to affect Tier-3, Tier-4 or deeper suppliers. The impact then propagates downstream via a bullwhip effect... In practice, that damping is often limited.

Continued on next page.

Table A.9 continued.

Code Name	Definition	Source Extract
Halted Flow Detection	The COVID-19 shock was immediately noticed by companies through stopped shipments, halted flows, and skyrocketing logistics costs.	They saw global transport prices skyrocket, or shipments simply stopped being delivered. So the problem manifested directly in halted flows and extreme logistics costs.
No Proactive Alert	There was no single internal system (dashboard, metric) that consistently flagged supply chain issues proactively.	There was no single proactive system or dashboard that consistently flagged issues.
AI Tier-N Risk Tool	A sophisticated AI-powered risk-analysis tool visualized tier-1 and tier-N risks, generating a prioritized management list, primarily for indirect procurement.	They used a supply chain risk-analysis tool that visualises tier-1 and tier-N risks. Using AI, it generates a prioritised list of risks that the company needs to manage. They mainly applied it to indirect procurement.
Conceptual Response	The interviewee's direct involvement was conceptual, focusing on designing governance and sourcing strategies, not implementing operational fixes.	I have not been involved in directly "fixing" a disruption at a client in an operational sense... this work was primarily conceptual: it focused on designing governance and approaches rather than implementing concrete mitigation actions.
Procurement Cost Focus	Procurement decisions are guided primarily by cost minimization, as the function struggles to articulate its strategic value related to resilience.	senior management largely perceived procurement as a cost-minimisation function... most procurement activities were framed in terms of reducing cost, rather than in terms of resilience, risk reduction, or strategic contribution.
VTT Resilience Pillars	The three foundational "must-haves" for resilience are Visibility, Transparency, and ultimately Traceability, which allows for active upstream risk reduction.	Resilience efforts... start with visibility and transparency... Moving from transparency to true traceability... With full traceability, you can then actively reduce upstream value-chain issues. So visibility, transparency, and ultimately traceability are the foundations of more resilient systems.

Continued on next page.

Table A.9 continued.

Code Name	Definition	Source Extract
Probabilistic Insights	Many resilience tools offer abstract, probabilistic insights (e.g., likelihood of a Tier-3 supplier) rather than providing concrete, traceable links or full chains of custody.	they often provide abstract, probabilistic insights rather than concrete, traceable links—for example, estimating the probability that a given company is your Tier-3 supplier without providing a full chain of custody.
Upstream Tool Focus	Most resilience tools concentrate on the upstream side (visibility/traceability), and few cover the full resilience cycle (identification and operational response).	Most tools were concentrated on the upstream side: visibility, transparency, or some form of traceability... very few that cover the full cycle. Typically, a tool is strong either in upstream visibility and risk identification, or in planning and response... but not both.
Data Quality Compensated	Tools actively compensate for fundamental data quality and integration challenges using external, open-source data and probabilistic estimation techniques.	Data quality and integration remain fundamental challenges. Many tools attempt to work around this by using open-source information to fill gaps or by leveraging whatever data is available and estimating the rest.
AI Deforestation Monitor	AI is used in solutions that combine geospatial data and image recognition to monitor supply chain activities, such as deforestation.	One example I find particularly compelling is a deforestation-monitoring solution... which uses geospatial data and AI-based image recognition to detect deforestation or potential deforestation activities.
Critical Data Sharing	Structured data sharing among partners is the most important collaboration mechanism for improving resilience by making deep-tier exposure more predictable.	Data sharing is, in my view, the most important collaboration mechanism... This requires integrated data and structured data sharing along the chain.
Competitive Data Reluctance	Companies are reluctant to share supplier data because they view it as a key component of their strategic competitive advantage, defining their market position.	many companies regard their supplier base as a key component of their strategic advantage, so they are understandably reluctant to share this information widely.

Continued on next page.

Table A.9 continued.

Code Name	Definition	Source Extract
Prioritize SCOR Source	The SCOR Source function is prioritized for copilot or agentic support because most supply chain disruptions originate in the upstream supplier base.	I would prioritise Source... Most disruptions originate in the upstream supplier base, so if that is where the risk lies, it makes sense to focus copilots and agentic support within Source first.

Themes and Supporting Open Coding Names

Table A.10: Themes and Supporting Open Coding Names - Second Interview (Supply Chain Expert)

Theme Name	Supporting Open Coding Name
Theme 1: Sources and nature of disruption	COVID Sourcing Shift; Regulatory Supplier Risk; Deep-Tier Bullwhip; Halted Flow Detection.
Theme 2: Resilience strategy and foundations	VTT Resilience Pillars; Conceptual Response; Procurement Cost Focus; SCM Strategy Overlay.
Theme 3: Technology for visibility and risk detection	AI Tier-N Risk Tool; No Proactive Alert; Upstream Tool Focus; Prioritize SCOR Source; AI Deforestation Monitor.
Theme 4: Data challenges and mitigation	Data Quality Compensated; Probabilistic Insights; Critical Data Sharing; Competitive Data Reluctance.

A.6. Interview 3 - Supply Chain Expert

Interview Summary

The interviewee works across all SCOR areas but focuses mainly on *Make* and *Deliver*. He sees key risks in localized network failures (e.g. flooding), deteriorating infrastructure that requires large investments, and increased traffic from online ordering. Problems are usually noticed only when stock is missing, in a reactive system where low inventories quickly amplify disruptions. To strengthen resilience, he recommends higher safety stocks, a ”traffic light” resilience strategy, classifying suppliers and customers by risk and using multi-sourcing for high-risk partners, together with clear strategies to recognize problems, align internally, and communicate externally.

He is skeptical of generic advanced AI tools due to poor data and missing context, and instead stresses the need for a robust IT system with continuous monitoring; AI is, in his view, useful mainly for pattern recognition under human supervision. Automation should start with simple, repetitive decisions and should never cover mission-critical tasks without human attendance. Finally, he emphasizes collaboration through clear communication, regular meetings, and treating suppliers and customers as cooperators, while honestly acknowledging that his own company can also be part of the problem.

Open coding

Table A.11: Open Coding Analysis – Third Interview (Supply Chain Expert)

Code Name	Definition	Source Extract
SCOR Focus (Make & Deliver)	The primary operational areas the interviewee concentrates on within the SCOR model.	If I must focus, I would position myself more on the make and delivery side.
Localized Network Failures	Specific, geographically contained breakdowns in infrastructure or delivery systems, often leading to temporary delays.	We have seen localized network failures, such as the flooding that broke down the railway and road network in Lindberg. On a smaller scale, I have mostly encountered delays of parcels.
Deteriorating Infrastructure Risk	The systemic challenge of aging infrastructure requiring massive financial investment, which impacts resilience.	On a systemic level, I worry about the quality of the infrastructure, which is deteriorating everywhere and needs billions of euros for maintenance and replacement.
Behavioral Traffic Risk (Internet Ordering)	Increased risk due to changes in consumer behavior, specifically the rise in traffic caused by online shopping.	Furthermore, I recognize the behavioral part as a key risk, as internet ordering creates more traffic.
Problem Notice via Missing Stock	Typical method for identifying a problem, which occurs when required inventory is physically unavailable for an order.	I usually see the problem when stock is needed for an order, and it is not in a rack. Staff then have to report that the item is on back order.
Reactive Issue Identification	Reliance on staff attention or customer complaints, meaning the system primarily responds to existing issues rather than anticipating them.	The issue can surface in two ways: either attentive personnel are aware, or the user has to complain. Since reliance on complaints means waiting for them, we are typically just responding (reactive).
Low Stock Amplification	Rapid escalation of disruptions when low inventory levels (maintained for cost reasons) cause missing parts to immediately halt production.	The disruption spreads because if I need every part to finish a product, any missing part is already blocking production. This is exacerbated when stocks are very low due to cost reasons.
Safety Stock Recommendation	Suggested mechanism of maintaining higher inventory levels to buffer against disruptions.	My proposed solution for resilience is to maintain additional safety stock or at least a higher safety stock level.

Continued on next page.

Table A.11 continued.

Code Name	Definition	Source Extract
Traffic Light Resilience System	Implementing a visual, categorized strategy (green, yellow, red) to guide operational responses under various conditions.	I believe a company should have a resilience strategy that uses a system like traffic lights (green, yellow, red) to guide actions under certain conditions.
Risk-Category Classification	Classifying supply chain partners (suppliers and customers) based on risk or behavior to tailor resilience strategies, including implementing multi-supply for risky suppliers.	I recommend classifying suppliers and customers by behavior, putting them into a risk category. For risky suppliers, the best practice is to always choose at least a second supplier (multi-supply).
Three Essential Resilience Steps	Three mandatory components for effective resilience implementation: strategy recognition, internal alignment, and clear external communication.	I believe the three essential steps a company must implement are: first, a strategy that allows me to recognize the problem; second, alignment on the internal solutions; and third, clear communication to the people outside.
Skepticism of Generic AI Tools	A cautious stance on advanced AI, favoring human judgment and citing common issues like lack of context and poor data quality in companies.	I personally do not use AI tools; I use my brain and believe it is still better. Many companies lack the correct data or the necessary context information.
Continuous Monitoring Function	A mandatory IT-system feature for resilience, designed to proactively alert users to potential problems (e.g., oversized orders).	My IT system needs a monitoring function that provides an alert if, for example, an order is way bigger than production capacity.
Multiple Disconnected IT Systems	Root cause of data quality and integration problems, resulting from various IT systems that are not properly linked.	Data quality issues arise because often companies have several IT systems that are not very well connected.
AI Pattern Recognition	AI provides clear value particularly in identifying recurring trends or anomalies.	The good functionality of AI is recognizing patterns.

Continued on next page.

Table A.11 continued.

Code Name	Definition	Source Extract
Gradual Automation of Simple Decisions	Recommended approach to automation: start with basic, repetitive tasks before moving to more complex ones.	I would start by allowing automated actions for the relatively simple decisions (the repetitive, basic ones). I would then gradually expand to more complex actions.
Guardrail: Avoiding Mission-Critical Automation	Safety rule: essential, high-stakes tasks must not be automated without human oversight, illustrated by a military failure case.	The most important guardrail is the awareness that some mission-critical things you should not do at least not unattended. I remember a military test where an automated cannon started shooting at the general's viewing area, which illustrates this risk.
Collaboration via Clear Communication	Foundational element for improving resilience through partnerships, characterized by clear channels, regular meetings, and mutual acceptance.	I believe it starts with good communication. I believe in open communication lines with clarity and also accepting certain things from each other.
Recognizing Shared Problem Ownership	Honesty in collaborative relationships: acknowledging that one's own company might also contribute to the supply chain problem.	I recommend viewing suppliers and customers as cooperators. I find that usually, when I dive into a problem, I am also part of it, and people need to be honest about that.

Themes and Supporting Open Coding Names

Table A.12: Themes and Supporting Open Coding Names - Third Interview (Supply Chain Expert)

Theme Name	Supporting Open Coding Name
Theme 1: Operational scope, disruption, and risk landscape	SCOR Focus (Make & Deliver); Localized Network Failures; Deteriorating Infrastructure Risk; Behavioral Traffic Risk (Internet Ordering).
Theme 2: Diagnosis and spread of issues	Problem Notice via Missing Stock; Reactive Issue Identification; Low Stock Amplification.
Theme 3: Strategic resilience and partnership	Safety Stock Recommendation; Traffic Light Resilience System; Risk-Category Classification; Three Essential Resilience Steps; Collaboration via Clear Communication; Recognizing Shared Problem Ownership.

Continued on next page.

Table A.12 continued.

Theme Name	Supporting Open Coding Name
Theme 4: Technology, AI, and automation guardrails	Skepticism of Generic AI Tools; Continuous Monitoring Function; Multiple Disconnected IT Systems; AI Pattern Recognition; Gradual Automation of Simple Decisions; Guardrail: Avoiding Mission-Critical Automation.

B

First Focus Group

B.1. First Focus Group

Open Coding Analysis

Table B.1: Open Coding Analysis – First Focus Group

Code Name	Definition	Source Extract
Pain Points and Process Challenges		
Infrequent-task inefficiency	People spend up to ten times longer on rarely occurring tasks due to lack of knowledge of correct procedures	Step 1 – Plan: “rare tasks take ∼10x longer without procedural knowledge”
Forecast unreliability	Client forecasts frequently fail to materialize as initially communicated	Step 1 – Source: “forecast shared, but demand does not materialise”
Customer-specific variability	The relationship is different for each customer, creating inconsistency	Step 1 – Source: “each customer requires a different way of working”
Waste reporting gaps	Product waste within hubs (expired/perished items) needs better reporting	Step 1 – Make: “waste in hubs needs systematic reporting”
Production risk monitoring	Dynamic monitoring of risk throughout the production process is needed	Step 1 – Make: “risk should be monitored continuously during production”

Continued on next page.

Table B.1 continued.

Code Name	Definition	Source Extract
Documentation burden	Significant effort required for documentation creation and coordination	Step 1 – Deliver: “documentation creation/coordination consumes major effort”
Fragmented delivery communication	Customer communication around deliveries is manual, fragmented, requires contacting multiple stakeholders, and leads to slow response times and errors	Step 1 – Deliver: “manual multi-stakeholder comms slows responses and increases errors”
Manual returns processing	Returns require manually retrieving information and investigating cases with hubs	Step 1 – Return: “returns require manual info retrieval and hub investigations”
Dashboard complexity	Numerous dashboards with relevant information, but navigating them requires prior knowledge and increases errors for new users	Step 1 – Enable: “many dashboards; hard to navigate without prior knowledge”
Supplier communication inconsistency	Different communication formats and practices are used for different suppliers	Step 1 – Enable: “suppliers use different formats and practices”
AI Agent Capabilities and Functions		
Sensor-based risk assessment	Agent gathers risk information using sensors tracking machine capacity	Steps 2&5: “use sensor/capacity signals to assess risk”
Automated report compilation	Agent compiles reports using live internal data, emails, and communication channels	Steps 2&5: “auto-compile reports from live data + emails”
Customer-facing chatbot	Agent interacts with customers via chatbot to capture feedback and process returns	Steps 2&5: “chatbot captures feedback and supports returns”
Expert knowledge digitization	Agent captures and provides access to knowledge held by few experts for infrequent tasks	Steps 2&5: “make expert-only knowledge accessible for rare tasks”
Demand signal filtering	Agent analyses historical orders and profiles to pre-filter credible vs. unreliable orders	Steps 2&5: “filter unreliable demand signals using history/profiles”

Continued on next page.

Table B.1 continued.

Code Name	Definition	Source Extract
Cross-dashboard aggregation	Agent aggregates and summarises information from multiple dashboards using semantic search	Steps 2&5: “semantic search to aggregate dashboards into one view”
Forecast deviation monitoring	Agent compares forecast changes with previous weeks and monitors stock levels	Steps 2&5: “track week-to-week forecast drift and stock impacts”
BBD-based waste detection	Agent checks best-before dates on randomized product samples to identify waste risk	Steps 2&5: “sample BBDs to detect emerging waste risk”
Automated returns data retrieval	Agent retrieves product information from unconnected databases for claims processing	Steps 2&5: “pull return/claim data across unconnected databases”
Governance, Permissions, and Boundaries		
Read-only machine access	Agent may retrieve machine information but never alter schedules	Step 3: “read machine info; never change schedules”
External communication prohibition	Agent must not send reports to or communicate with external parties	Step 3: “no external comms without human authorisation”
Financial transaction restriction	Agent cannot perform automatic transactions or access bank information	Step 3: “no banking access; no automatic transactions”
Owner consent for confidential data	Agent cannot access confidential information without the explicit consent of its owner	Step 3: “confidential data requires explicit owner consent”
Human approval for decisions	Agent must not take direct decisions (e.g., modifying delivery components) without prior human approval	Step 3: “recommendations ok; decisions require approval”
Row-level security scoping	Data access controlled via row-level security to prevent scope breach	Step 3: “row-level controls prevent scope creep”
Supplier data sovereignty	Agent must not access supplier systems unless explicitly shared; suppliers may not want AI handling their data	Step 3: “supplier systems only if explicitly shared”

Continued on next page.

Table B.1 continued.

Code Name	Definition	Source Extract
Mandatory human supervision	Agent must not fully manage systems without human oversight	Step 3: “no full autonomy; humans supervise”
Information immutability	Agent may store but not share, manage, or modify stored information	Step 3: “store context; do not redistribute or modify”
Rollback and Escalation Patterns		
Human-in-the-loop verification	Human operator reviews AI outputs for consistency with real conditions (e.g., sensor readings, production data)	Step 4: “human checks outputs against operational reality”
Content validation and correction	Human reviewer validates, corrects, or removes misleading or confidential content from AI reports	Step 4: “validate/correct; remove misleading or confidential content”
Handover to human support	Complex or inaccurate cases are handed to human agents for resolution	Step 4: “handoff when complexity/accuracy exceeds scope”
Supervisory second agent	A second AI agent monitors the primary agent’s data sources and can restart workflows	Step 4: “supervisor agent monitors and can restart”
Infinite loop detection	After $\sim$1,000 iterations in a loop, a human operator is notified to intervene	Step 4: “loop threshold triggers human intervention”
Hard-coded process control	A hard-coded filter requires both a suggested process and explicit human approval; mitigates prompt injection	Step 4: “hard-coded gates + approval mitigate prompt injection”
Flawed forecast correction	A flawed statistical forecast triggers manual intervention to correct supply chain planning	Step 4: “forecast failure triggers manual correction”
Outlier detection agent	A second AI agent identifies outliers and restarts algorithms for suspected items	Step 4: “detect outliers; restart when anomalies suspected”

Continued on next page.

Table B.1 continued.

Code Name	Definition	Source Extract
Data integrity fallback	Secondary fixed dashboard with verified but non-live data as fallback when live data is incorrect	Step 4: “fallback to verified reference dashboard when live is wrong”

Themes and Supporting Open Coding Names

Table B.2: Themes and Supporting Open Coding Names – First Focus Group

Theme Name	Associated Codes
Information Fragmentation Under Load	Dashboard complexity; Fragmented delivery communication; Supplier communication inconsistency; Documentation burden; Manual returns processing; Cross-dashboard aggregation; Automated report compilation; Automated returns data retrieval
Procedural Rigidity in Exception Scenarios	Infrequent-task inefficiency; Expert knowledge digitization; Dashboard complexity; Human approval for decisions
Detection Lag in Slow-Moving Risks	Production risk monitoring; Waste reporting gaps; Sensor-based risk assessment; BBD-based waste detection; Forecast deviation monitoring; Forecast unreliability
Feedback Loop Atrophy	Documentation burden; Manual returns processing; Customer-facing chatbot; Automated report compilation; Content validation and correction; Handover to human support
Diagnostic & Monitoring Agents	Sensor-based risk assessment; Production risk monitoring; Waste reporting gaps; BBD-based waste detection; Forecast deviation monitoring; Outlier detection agent
Coordination & Evidence-Building Agents	Cross-dashboard aggregation; Automated report compilation; Automated returns data retrieval; Manual returns processing; Documentation burden; Fragmented delivery communication
Interface & Learning Agents	Expert knowledge digitization; Infrequent-task inefficiency; Customer-facing chatbot; Demand signal filtering; Handover to human support
Always Allowed	Read-only machine access; Cross-dashboard aggregation; Automated report compilation; Automated returns data retrieval; Information immutability
Never Allowed	External communication prohibition; Financial transaction restriction; Human approval for decisions; Mandatory human supervision; Supplier data sovereignty; Row-level security scoping; Owner consent for confidential data
Human Validation	Human-in-the-loop verification; Content validation and correction; Handover to human support

Continued on next page.

Table B.2 continued.

Theme Name	Associated Codes
Secondary Agent Supervision	Supervisory second agent; Infinite loop detection; Hard-coded process control
Data Quality Controls	Outlier detection agent; Data integrity fallback; Content validation and correction
Escalation Thresholds	Handover to human support; Infinite loop detection; Flawed forecast correction; Human-in-the-loop verification

Agents Portfolio

Table B.3: Agents derived from focus group

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Capacity and risk assessment Agent	Monitors machine-capacity risk using sensors and internal/external signals; informs production manager about ability to meet demand; escalates when capacity limits/bottlenecks are reached.	Retrieve information about machines in use.	Make changes to schedules.	*Rollback:* Human operator reviews capacity outputs and verifies sensor readings if miscalculations occur. *Escalation:* Notify machine operators to check sensors; escalate to production management if inconsistencies persist or performance is impacted.
Documentation & reporting Agent	Compiles end-to-end supply chain reports from live internal data and communications within agreed limits; informs stakeholders; escalates if the report is inaccurate.	Change format of attached documents and files.	Automatically send reports or communicate outside the company.	*Rollback:* Human reviewer validates output, removes/corrects sensitive or inaccurate content; adjust/re-train AI if needed. *Escalation:* To compliance/data protection for confidentiality issues; to business owners/management for misleading or high-impact insights (content-dependent).

Continued on next page.

Table B.3 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Feedback Agent	Collects customer feedback via chatbot; schedules returns, initiates reimbursements, prepares documentation; informs process owner; routes high-value transactions to human verification.	Access invoices and client information.	Perform automatic transactions or access bank information.	*Rollback:* Handover to human customer support for complex/inaccurate cases; agent resolves manually and requests clarifications if needed. *Escalation:* To client support via an explicit handover mechanism (e.g., dedicated button).
Expert process guidance Agent	Supports rare/complex expert tasks via predefined prompts/templates; captures relevant information; informs stakeholders; escalates to IT if the process is missing or fails.	Access confidential, rarely used expert information on demand.	Access that information without explicit consent of its owner.	*Rollback:* Second supervisory agent audits data sources during retrieval and can restart workflow. *Escalation:* If an infinite loop occurs (e.g., repeated confidential retrieval + continuous triggers), notify a human operator after a threshold (e.g., 1,000 iterations) to exclude data from the workflow.
Demand forecast support Agent	Analyses historical orders, profiles, and sales feedback to pre-filter incoming requests; flags credible vs. unreliable demand; aligns with Sales; recommends planning scenario to Supply Chain; escalates on disruption signals or recurring forecast issues.	Ask questions about customer needs, access current customer data, and act as an employee when reaching out.	Take direct decisions (e.g., modify delivery components) without prior human approval.	*Rollback:* Hard-coded control gate requires both (i) suggested action and (ii) explicit human approval; otherwise process stops (mitigates prompt-injection). *Escalation:* If the hard-coded gate blocks too frequently (e.g., repeated customer requests), trigger human intervention.

Continued on next page.

Table B.3 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Insights Agent	Aggregates and summarises information across dashboards; highlights relevant metrics and procedures to change them; uses semantic search across KPI definitions and usage patterns; informs process owner/ops manager; escalates when data are missing/inconsistent.	Use SQL server for structured collection and enable RAG across multiple sources beyond dashboards.	Access confidential data outside defined scope (enforceable via row-level security).	*Rollback:* Second agent monitors data sources during retrieval and can restart workflow. *Escalation:* If infinite loop arises (e.g., repeated confidential retrieval + continuous triggers), notify a human operator after a threshold (e.g., 1,000 iterations) to remove those data from the workflow.
Forecast monitoring Agent	Compares updated forecasts against prior weeks within thresholds and monitors supplier stock signals; informs Demand Planning; escalates on large deviations or low stock.	Access internal planning data and historical path information.	Directly access supplier availability/systems unless explicitly shared.	*Rollback:* If statistical outputs are flawed and could drive wrong production decisions, revert to manual review/correction. *Escalation:* SNH manually intervenes to correct supply chain planning.
Waste monitoring Agent	Checks best-before dates (BBD) on random product samples; combines BBD, shelf-life history, and prior waste levels; informs hub managers; escalates to Supply Chain if many items risk perishing.	Store and automatically collect relevant data from necessary sources.	Fully manage the system without human supervision.	*Rollback:* If incorrect BBDs disrupt the algorithm, trigger remediation. *Escalation:* Secondary AI detects outliers and restarts algorithm for suspected items; if persistent, escalate to a human.

Continued on next page.

Table B.3 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Claims Agent	Retrieves key product/-claim data (SKU, order number, batch/lot) by querying currently unconnected databases; informs suppliers directly; escalates for large volumes or high-value returns.	Store all relevant information.	Share, manage, or modify this information.	*Rollback:* If incorrect data are gathered (e.g., SKU mix-ups, date-format errors), IT investigates/corrects data—optionally supported by a secondary fixed dashboard with verified (non-live) information. *Escalation:* Escalate to IT promptly; escalate also when high item counts or high total return value are detected.

B.2. Follow-up Interview

Open Coding

Table B.4: Open Coding Analysis – Follow Up Interview

Code Name	Definition	Source Extract
Extra step burden	The claims process introduces an additional step for operators, which discourages them from completing tickets.	”The process, however, requires an extra step: the operator has to decide what to send out and at the same time explain the reason in the ticket.”
Undeclared waste	Defective items are not properly recorded, leading to hidden or undeclared waste in the system.	”They often skip it, and this leads to a lot of undeclared waste.”
Misleading supplier performance	Missing defect registration makes supplier performance appear worse or different from reality.	”It can even look as if the supplier never delivered a certain item, when in reality it was delivered but later flagged as defective.”
Loss of visibility	Failure to record defects erodes visibility on inventory quality and flows.	”When operators fail to correctly record defective items, visibility is lost: stock discrepancies appear, waste remains undeclared.”

Continued on next page.

Table B.4 continued.

Code Name	Definition	Source Extract
Feedback loop breakdown	Lack of timely feedback to hubs leads them to stop reporting problems.	”Over time they stop reporting because they feel nothing happens after they send a ticket.”
Communication constraints	Operational workload and limited screen time hinder rich, two-way communication.	”They’re extremely busy and not sitting at a computer waiting for our messages, so communication becomes fragmented and slow.”
Unclear ticket requirements	Operators are uncertain about which information must be included in tickets.	”People were often unsure which information they were supposed to enter in the ticket.”
Desire for real time reporting	Operators want to report defects while doing inbounding rather than as a later, separate task.	”They preferred to report issues while they were actually doing inbounding, rather than stopping and resuming later.”
Process standardisation benefit	The new internal app standardises and simplifies ticket creation for all operators.	”This redesign made the process more standardised, simpler and faster, so that any operator can handle it without particular training.”
Standardisation information loss	Standardised categories reduce the richness and nuance of free-text explanations.	”Standardisation reduces space for free-form comments... the richer descriptive comments that were available before appear less frequently.”
Ticket volume overload	The easier process results in a surge of tickets that cannot be handled manually.	”On a single day there were already around ninety-six tickets, and at that point it becomes impossible to manage them one by one as before.”
Accountability assignment need	There is a need to determine who is responsible for each defect (supplier vs. logistics).	”The task is not limited to understanding what the problem is but also to determining who is accountable for it.”
Time pressure on claims	Strict supplier deadlines for reporting defects create strong time pressure.	”Many suppliers also impose strict rules, such as reimbursing damaged goods only if the issue is reported within forty-eight hours.”
AI for speed and scale	AI is adopted primarily to cope with volume and speed requirements.	”Either you spend the whole day analysing tickets or you automate it... AI is therefore necessary to achieve the required speed and scale.”
Grouping before reasoning	The agent first groups and structures tickets before deeper analysis.	”From a single stream of tickets, the system organises cases according to the type of problem and the party that appears accountable.”

Continued on next page.

Table B.4 continued.

Code Name	Definition	Source Extract
AI email drafting	The agent reformulates comments and data into ready-to-send claim emails.	”The system reads the operator’s comment... and then drafts an email that is ready, or almost ready, to be sent.”
Legacy flow richer data	The old, less standardised system provides richer input for the advanced agent.	”This flow was designed on the basis of the old system, in which operators wrote complete free-text messages, so the underlying information is richer.”
Future flow risk	The more advanced agentic flow is seen as a future, somewhat risky version.	”There is currently a smaller and safer version for everyday use, and a more ambitious future version that is still somewhat risky.”
Human escalation criteria	Ambiguous or edge cases are escalated to a human for review.	”In such cases the system escalates the issue to a human... the case is reviewed manually.”
Supervisory human role	The human role shifts from execution to supervision and final checking.	”The human role becomes more supervisory and focused on final checks... Over time this reduces repetitive manual tasks.”
Manual routine as blueprint	The manual workflow serves as the direct blueprint for the automated n8n flow.	”The work consisted of always the same sequence... This sequence was translated almost directly into an automated flow.”
Limited value variance	Similar item values reduce the usefulness of value-based automation thresholds.	”The value of individual items is usually quite similar... introducing a threshold... would not dramatically change the workload.”
Agentic decision orientation	The use case emphasises decision-making and orchestration rather than pure text generation.	”The core requirement is not to produce text... but to take decisions, assign accountability, route each case to the right party and orchestrate a sequence of operations.”
GenAI for email only	Generative AI is considered mainly for formatting and wording emails.	”If the only goal were to draft messages, a generative model would be enough.”

Themes and Supporting Open Coding Names

Table B.5: Themes and Supporting Open Coding Names – Follow Up Interview

Theme Name	Associated Codes
Theme 1: Reporting friction and hidden waste	Extra step burden; Undeclared waste; Misleading supplier performance; Loss of visibility; Feedback loop breakdown; Communication constraints
Theme 2: Operator needs and process redesign	Unclear ticket requirements; Desire for real time reporting; Manual routine as blueprint
Theme 3: Standardisation trade-offs in data quality	Process standardisation benefit; Standardisation information loss; Legacy flow richer data
Theme 4: Scale, time pressure and rationale for AI adoption	Ticket volume overload; Time pressure on claims; AI for speed and scale; Limited value variance
Theme 5: Two-step agent workflow and evolving capabilities	Grouping before reasoning; AI email drafting; Future flow risk; Legacy flow richer data
Theme 6: Human-in-the-loop supervision and escalation logic	Human escalation criteria; Supervisory human role; Communication constraints
Theme 7: Agentic vs. generative AI positioning	Agentic decision orientation; GenAI for email only; AI email drafting

Themes and Definitions

Table B.6: Themes and Definitions – Follow Up Interview

Theme Name	Definition
Theme 1: Reporting friction and hidden waste	The legacy claims workflow introduces additional cognitive and procedural effort for operators, which discourages complete reporting of defects and leads to systematic under-documentation of waste and supplier performance issues.
Theme 2: Operator needs and process redesign	The redesigned process emerges from a systematic diagnosis of operator pain points, revealing the need for clearer information requirements and for defect reporting to be integrated directly into real-time inbounding activities.

Continued on next page.

Table B.6 continued.

Theme Name	Definition
Theme 3: Standardisation trade-offs in data quality	Internal app standardisation simplifies and speeds up ticket creation for all operators, but at the cost of losing rich free-text context, thereby constraining what downstream analytical and AI components can infer from the data.
Theme 4: Scale, time pressure and rationale for AI adoption	Growing ticket volumes and strict supplier deadlines create a structural need for automation, positioning AI as the only viable way to handle claims at the required speed and scale without disproportionate human workload.
Theme 5: Two-step agent workflow and evolving capabilities	The agent is conceptualised as a two-phase system that first structures and groups claims, then synthesises comments and contextual data into ready-to-send emails, with a gradual roadmap from basic automation to more advanced image and text understanding.
Theme 6: Human-in-the-loop supervision and escalation logic	Despite automation, humans remain central as final decision-makers and exception handlers, with explicit escalation criteria and a shift from manual execution toward supervisory control and oversight of the agent's outputs.
Theme 7: Agentic vs. generative AI positioning	The use case is framed primarily as an agentic decision and orchestration problem, where generative AI plays a limited, supportive role in drafting textual outputs rather than driving core reasoning or accountability decisions.

C

Second Interview Iteration

C.1. Interview 1 - Supply Chain Expert

Interview Summary

The interviewee assessed the proposed agent portfolio and determined that, while individual agents address specific issues such as documentation gaps, demand volatility, waste, and claims, none comprehensively manages disruptions at the broader network level. Key areas for improvement include improved IT integration, greater domain specialization, and more effective translation of monitoring signals into actionable measures.

An additional agent was proposed: a regulatory and constraint monitoring agent that tracks external compliance signals, such as trade restrictions, regional shutdowns, and policy changes, and converts them into actionable supply chain reconfigurations. Ideally, this agent would be operated by a neutral third party. Regarding human oversight, the interviewee emphasized that the appropriate level of autonomy is domain-dependent. Safety-critical contexts, such as pharmaceuticals, require mandatory human sign-off, whereas low-impact and easily reversible environments can accommodate higher levels of automation. The Part 2 framework was considered strong and comprehensive, with only a minor classification question regarding whether ”routines” should be categorized under ”structural levers.”

Open Coding

Table C.1: Open Coding Analysis – First Interview (Supply Chain Expert)

Code Name	Definition	Source Extract
Domain dependence of AI applications	The usefulness and design of AI agents in supply chains depend on the specific function, domain, and decision context in which they are deployed.	“How are you modelling the supply chain? Have you identified fundamental functions like purchasing? Doesn’t its application depend a bit on the specific domain?”

Continued on next page.

Table C.1 continued.

Code Name	Definition	Source Extract
Different uncertainty levels across decisions	Different supply chain decisions are exposed to different degrees and types of uncertainty, which affects how suitable prediction-based AI support can be.	"Different decisions in a supply chain deal with different levels of uncertainty."
Limits of historical-data forecasting	Forecasting approaches based only on historical data are limited because they assume system stability and may fail under structural change or major disruptions.	"If you have a good amount of data, you can make predictions assuming the system won't change. But in the long term, a model might not anticipate a massive, unforeseen disruption."
Data collection as resilience prerequisite	Robust data collection is a foundational requirement for meaningful AI-based resilience support and is increasingly recognized by companies as essential.	"The data collection aspect is crucial—companies are finally realising that this is essential, whereas 10 years ago it wasn't seen as a problem."
Capacity management reduces uncertainty	A capacity-monitoring agent primarily supports uncertainty reduction by making operational capacity more visible and manageable.	"I would say uncertainty, as it's a capacity management tool."
Human-in-the-loop for capacity decisions	Capacity-related AI support should not be fully autonomous; it should complement operational oversight rather than replace human judgement.	"Dynamic task allocation and capacity management, while keeping the operator involved."
Maintenance linkage from operational data	Operational utilization data can be extended beyond monitoring to support predictive or planned maintenance decisions.	"I would connect it to maintenance. Machine utilisation provides information about the equipment's life cycle."
Machine maintenance agent opportunity	The same operational data used for capacity monitoring can enable additional agentic applications such as maintenance scheduling.	"A machine maintenance agent. It could use the same data source to program maintenance schedules."
Bottleneck and demand load visibility	Monitoring machine loading over time can support bottleneck identification and broader demand management decisions.	"Maybe a bottleneck analysis or demand management agent, to see which machine is the most heavily loaded over time."
Information as uncertainty management resource	Access to timely and relevant information is central to managing uncertainty in supply chains.	"I would give it a 1 for risk, and a 2 for uncertainty, as information is key to managing uncertainty."

Continued on next page.

Table C.1 continued.

Code Name	Definition	Source Extract
Information gaps affect quality	A lack of structured information undermines both process quality and product quality, making reporting agents valuable.	"It solves the lack of information, improving both product and process quality."
Need for report specificity	Broad reporting agents become more useful when tailored to specific subcategories and operational purposes.	"I would make it more specific by tailoring it to various subcategories."
Quality management automation opportunity	Documentation and reporting capabilities can be extended into automated quality management and traceability applications.	"A quality management agent. It could automatically document quality control inspections and link that data to the product's history."
Supplier capability documentation for reconfiguration	Documenting supplier capabilities can support supply chain reconfiguration and improve adaptive resilience.	"An agent that manages supplier documentation and their functional capabilities could be a great way to help reconfigure the supply chain."
Customer feedback as resilience input	Customer feedback is an important resilience signal because it captures market response and downstream operational issues.	"Feedback is fundamental for resilience."
Feedback transforms risk into operations	Feedback-related agents help translate latent risks into operationally actionable disturbances.	"I would say it impacts risk and disturbances, transforming a risk into something operational."
Reverse-flow streamlining	Feedback agents help streamline reverse flows such as returns, reimbursements, and information transfer from customers back into the supply chain.	"It streamlines the flows from the customer back to the supply chain."
Customer sentiment as demand signal	Customer sentiment is a relevant source of information because it directly influences demand formation and response.	"It captures customer sentiment, which is ultimately what generates demand."
Distributed feedback across supply chain actors	Feedback mechanisms should not be limited to end customers but should be extended to all actors across the supply chain.	"I would expand it so that every element of the supply chain can provide feedback, not just the end customer."
Circularity and end-of-life management	AI opportunities exist in linking feedback and returns to circularity-oriented product end-of-life management.	"Something focused on circularity and the end-of-life management of products."

Continued on next page.

Table C.1 continued.

Code Name	Definition	Source Extract
Expert guidance reduces process errors	Operationalising expert knowledge through AI guidance can reduce errors and support more reliable task execution.	"It prevents errors and improves both process and product quality."
Knowledge support for changing demand	Expert guidance systems could become more valuable if they adapt instructions proactively when task requirements shift due to demand changes.	"If tasks change because demand changes, the agent could proactively predict and suggest new instructions based on those changes."
Skill-based resource allocation	Knowledge-intensive processes create opportunities for agents that allocate resources and tasks according to required skill profiles.	"An agent for resource or task allocation based on the specific skills required for those complex processes."
Demand forecast support targets disturbances	Forecast-support agents are most useful in mitigating disturbances by comparing incoming demand with historical patterns to improve planning.	"I'd say it works on mitigating disturbances. I'll give it a 2 for disturbance."
Dashboard consolidation for decision support	Consolidating large volumes of dispersed operational information can improve decision speed and quality.	"Imagine a person who has to check 50,000 dashboards; this agent consolidates that information so they can make correct decisions quickly."
Insight agents for bottleneck resolution	Insight-oriented agents support operational improvement by identifying bottlenecks and improving process visibility.	"Resolving bottlenecks and general process improvement."
Forecast monitoring as event-based disturbance response	Monitoring forecast deviations is primarily an event-driven activity aimed at responding to disturbances rather than eliminating uncertainty.	"It looks at changes in demand and flags them, acting on an event. So it acts on disturbances."
Demand forecasting and purchasing interdependence	Forecast monitoring should be connected to purchasing because demand signals and material acquisition are strongly interdependent.	"I would connect it to the purchasing management agent. Demand forecasting and purchasing materials are heavily intertwined."
Waste monitoring for proactive risk management	Waste-monitoring agents are valuable because they act before products perish, reducing risk and partially addressing disturbances.	"It acts on upcoming risks. If it has full product history, I'd give it a 2 for risk and a 1 for disturbance."

Continued on next page.

Table C.1 continued.

Code Name	Definition	Source Extract
Obsolescence prevention	Waste-monitoring logic supports resilience by preventing product obsolescence and enabling timely intervention.	"Preventing product obsolescence and managing unexpected events."
Fast-routing linkage for at-risk goods	Monitoring perishability or waste becomes more effective when connected to execution agents that can rapidly reroute goods.	"I would connect it to an agent that manages fast routing for delivery, to dispatch the product quickly once it's identified as at risk."
Supplier-side risk prevention through visibility	Supplier-side agents help prevent disruptions by consolidating fragmented information and improving visibility on supplier-related risks.	"It prevents supply disruptions by ensuring you are informed about potential supplier failures."
Order management and delay prevention	Claims or supplier-support agents contribute to resilience by improving order management and reducing the likelihood of material delays.	"Order management and preventing delays in material delivery."
System integration as enabling condition	The effectiveness of supplier-oriented AI agents depends heavily on strong IT integration across systems, devices, and platforms.	"It relies heavily on IT system integration—making different devices and platforms talk to each other."
Traceability for circular materials management	There is potential for agents that trace critical materials across the supply chain to support circularity and end-of-life recovery.	"Tracking rare metals like magnets in electric scooters, from extraction in China all the way to their end-of-life in Europe."
Regulatory and policy monitoring as resilience function	A valuable resilience application is a high-level agent that monitors changing regulations and policies affecting supply chain design and operations.	"A high-level agent that monitors regulations and policies."
Inter-organisational communication gaps	Supply chains often underperform because firms do not communicate effectively across organisational boundaries.	"Supply chains often struggle because individual companies don't communicate well with each other."
Compliance-driven reconfiguration	AI can support resilient reconfiguration by monitoring regulatory shifts and helping firms adapt supply chains across regions.	"An agent could monitor environmental regulations, emissions policies, or geopolitical closures, and help reconfigure the supply chain to remain compliant and resilient."

Continued on next page.

Table C.1 continued.

Code Name	Definition	Source Extract
Human oversight for safety-critical functions	Safety-critical or high-quality-impact decisions should retain human oversight, especially where failure consequences are severe.	"Anything with an impact on safety or critical quality—like in the pharmaceutical industry—requires a human in the loop."
Humans as flexible safeguard	Humans remain the most flexible and reliable safeguard for managing high-risk functions when automation cannot be guaranteed to be fail-safe.	"Unless you can be 100% certain the automation won't fail, humans are the most flexible resource you can use for managing high-risk functions."

Themes and Supporting Open Coding Names

Table C.2: Themes and Supporting Open Coding Names – First Interview (Supply Chain Expert)

Theme Name	Associated Codes
Theme 1: Context Dependence of AI Deployment in Supply Chains	Domain dependence of AI applications; Different uncertainty levels across decisions; Limits of historical-data forecasting; Data collection as resilience prerequisite
Theme 2: Monitoring and Visibility as Core Enablers of Resilience	Capacity management reduces uncertainty; Information as uncertainty management resource; Dashboard consolidation for decision support; Insight agents for bottleneck resolution; Supplier-side risk prevention through visibility; Order management and delay prevention
Theme 3: Human-Centred AI for Operational Control	Human-in-the-loop for capacity decisions; Expert guidance reduces process errors; Human oversight for safety-critical functions; Humans as flexible safeguard
Theme 4: AI as a Driver of Operational Improvement and Adaptive Coordination	Maintenance linkage from operational data; Machine maintenance agent opportunity; Bottleneck and demand load visibility; Need for report specificity; Quality management automation opportunity; Supplier capability documentation for reconfiguration; Knowledge support for changing demand; Skill-based resource allocation; Demand forecasting and purchasing interdependence; Fast-routing linkage for at-risk goods; System integration as enabling condition
Theme 5: Feedback, Traceability, and Circularity as Extended Resilience Functions	Customer feedback as resilience input; Feedback transforms risk into operations; Reverse-flow streamlining; Customer sentiment as demand signal; Distributed feedback across supply chain actors; Circularity and end-of-life management; Traceability for circular materials management
Theme 6: Disturbance and Risk Mitigation Through Event-Based Agents	Demand forecast support targets disturbances; Forecast monitoring as event-based disturbance response; Waste monitoring for proactive risk management; Obsolescence prevention

Continued on next page.

Table C.2 continued.

Theme Name	Associated Codes
Theme 7: External Awareness and Compliance-Oriented Reconfiguration	Regulatory and policy monitoring as resilience function; Inter-organisational communication gaps; Compliance-driven reconfiguration

Thematic Analysis

Table C.3: Themes and Definitions – First Interview (Supply Chain Expert)

Theme Name	Definition
Theme 1: Context Dependence of AI Deployment in Supply Chains	This theme captures the interviewee's view that the value of AI in supply chains is inherently context-dependent. The relevance, design, and expected performance of AI agents vary according to the specific supply chain domain, the function being supported, and the level of uncertainty surrounding the decision. The interviewee stresses that different decisions are exposed to different uncertainty conditions, and that prediction-based support is therefore not uniformly reliable across contexts. In particular, approaches based only on historical data are seen as limited because they assume relative system stability and may fail when structural shifts or large disruptions occur. As a result, effective AI deployment requires not only contextual adaptation but also strong underlying data collection practices, which are framed as a foundational prerequisite for any meaningful resilience-oriented application.
Theme 2: Monitoring and Visibility as Core Enablers of Resilience	This theme reflects the central importance attributed to monitoring, reporting, and visibility in strengthening supply chain resilience. Across several agent examples, the interviewee consistently links resilience value to the ability to make operational conditions, risks, and emerging problems more visible. Capacity monitoring reduces uncertainty by clarifying available resources, reporting agents reduce information gaps, and insight agents help decision-makers navigate large volumes of fragmented information. Likewise, supplier-side visibility is viewed as essential for identifying potential failures early and preventing downstream disruption. Overall, the theme suggests that one of the main contributions of AI is not necessarily autonomous decision-making, but the creation of better informational awareness through which organisations can identify bottlenecks, reduce ambiguity, and improve decision quality.
Theme 3: Human-Centred AI for Operational Control	This theme highlights the interviewee's preference for AI systems that support, rather than replace, human decision-makers in operational environments. Even when AI agents provide meaningful assistance in capacity management, process guidance, or risk detection, the interviewee emphasises the continued need for human involvement, especially where tasks affect safety, critical quality, or high-risk operations. Humans are described as the most flexible and reliable safeguard when automation cannot be guaranteed to be fail-safe. The theme therefore frames resilience-oriented AI as fundamentally human-centred: AI may improve consistency, reduce errors, and structure knowledge, but final control should remain with humans whenever operational consequences are severe or uncertainty cannot be fully contained.

Continued on next page.

Table C.3 continued.

Theme Name	Definition
Theme 4: AI as a Driver of Operational Improvement and Adaptive Coordination	This theme captures the interviewee's view that AI can move beyond passive monitoring to enable more proactive and coordinated operational improvements. Several proposed extensions illustrate how data and insights generated in one area can be linked to broader decision functions, such as maintenance scheduling, bottleneck analysis, quality management, skill-based allocation, purchasing coordination, and fast-routing responses. The recurring logic is that resilience is strengthened when isolated monitoring functions are integrated into larger adaptive coordination mechanisms. AI agents become particularly valuable when they connect different parts of the supply chain, align information across functions, and support operational reconfiguration in response to evolving conditions. This theme therefore reflects a systemic view of resilience in which AI supports coordination across processes rather than optimising individual tasks in isolation.
Theme 5: Feedback, Traceability, and Circularity as Extended Resilience Functions	This theme concerns the expansion of resilience thinking beyond internal operations toward downstream feedback, reverse flows, and end-of-life traceability. The interviewee views customer feedback as a critical resilience input because it transforms latent problems into operational signals and reveals how products and services are experienced in practice. At the same time, feedback is not limited to customers alone; the interviewee suggests that all actors in the supply chain should contribute to this informational loop. From this perspective, AI can support resilience not only by managing returns and customer sentiment, but also by enabling broader forms of traceability and circularity, such as tracking materials and supporting product end-of-life management. The theme thus broadens resilience from disruption response to include learning, recovery, and resource recirculation across the product lifecycle.
Theme 6: Disturbance and Risk Mitigation Through Event-Based Agents	This theme reflects the interviewee's interpretation of many AI agents as tools for identifying, flagging, and responding to concrete operational events rather than eliminating uncertainty at its source. Demand forecast support, forecast monitoring, and waste monitoring are all described as agents that operate on disturbances or upcoming risks by detecting deviations, flagging critical changes, and enabling timely responses. Their value lies in transforming emerging problems into actionable signals before they escalate further. This theme suggests that a significant portion of resilience-oriented AI is event-based in nature: it focuses on monitoring operational variation, identifying material deviations from expectations, and supporting interventions that reduce the impact of risk and disturbance on system performance.

Continued on next page.

Table C.3 continued.

Theme Name	Definition
Theme 7: External Awareness and Compliance-Oriented Reconfiguration	This theme captures the interviewee's emphasis on resilience as a challenge shaped not only by internal operations but also by the broader regulatory, geopolitical, and inter-organisational environment. The interviewee identifies policy and regulation monitoring as a particularly valuable high-level application of AI, especially in contexts where environmental requirements, geopolitical closures, or regional compliance conditions force supply chains to adapt. At the same time, the theme highlights the persistent difficulty of communication across organisational boundaries, which limits supply chain coordination and reconfiguration. AI is therefore seen as potentially valuable in helping firms interpret external change, maintain compliance across regions, and redesign supply chain structures in response to evolving external constraints. In this sense, resilience is framed not only as operational recovery, but also as the capability to remain viable under shifting institutional and geopolitical conditions.

C.2. Interview 2 - Supply Chain Expert

Interview Summary

The interviewee evaluated the agent portfolio as most effective in disturbance management and risk monitoring, but noted limited effectiveness in uncertainty reduction and managing full network disruptions. Two overarching improvement themes were identified: agents should transition from passive monitoring to proactive, action-oriented operations, and should be integrated across the entire process rather than functioning in silos. At the individual agent level, recommended enhancements include implementing simulation capabilities for capacity planning, improving master-data integration for documentation, adopting scenario-based forecasting for demand support, and utilizing predictive claim pattern recognition for the claims agent. It was also suggested that the feedback agent be divided into a dedicated analysis agent and an operational handling agent.

Further opportunities for generative AI were identified in network design, supply-demand balancing, customer service, and data quality improvement. Regarding human oversight, the interviewee emphasized the necessity of clear acceptance criteria, flagging rules, and defined operational boundaries, as AI outputs may appear plausible but lack accuracy. The existing guidelines were regarded as coherent and robust, with a recommendation to add an explicit guideline on cross-team harmonization and alignment with formal organization-wide sign-off procedures. Finally, the importance of embedding resilience agents within coordinated team processes, rather than deploying them informally or in isolation, was underscored.

Open Coding

Table C.4: Open Coding Analysis – Second Interview (Supply Chain Expert)

Code Name	Definition	Source Extract
Anonymity and non-client framing	Expert input should be framed as personal professional judgement rather than as client-specific or company-specific evidence.	"It is only my personal opinion based on my experience and what I have seen. There will be no connection to any clients... it is simply my view."

Continued on next page.

Table C.4 continued.

Code Name	Definition	Source Extract
AI as support for summarization	AI is seen as particularly useful for summarization and research support tasks.	“I guess AI would also help a lot with the summaries.”
Interrelated resilience dimensions	Uncertainty, risk, disturbance, and disruption are understood as interrelated rather than fully separable categories.	“They are normally interrelated, so it is good that we can assign more than zero to multiple categories.”
Disturbance-oriented monitoring value	Capacity monitoring creates resilience value primarily by making operational disturbances visible once thresholds are exceeded.	“I would say disturbance would be the most relevant category... whenever you exceed something, it becomes visible.”
Historical monitoring supports risk detection	Tracking performance over time allows monitoring agents to identify growing risk trends before they escalate.	“I would give a one to risk, because if someone is monitoring, they can also see whether a problem has a growing trend by checking historical data.”
Predictive maintenance support	Monitoring agents help identify early signals relevant for predictive maintenance and machine breakdown prevention.	“This monitoring aspect helps a lot with machine breakdowns and predictive maintenance.”
KPI-based operational tracking	Operational resilience agents rely on specific KPIs such as vibration, speed, or handled volume to track system health.	“This agent identifies different KPIs you have set—like machine vibrations, speed limits, or volume handling.”
Bottleneck identification through flow visibility	Monitoring operational volumes enables bottleneck identification across warehouses, consolidation points, or networks.	“Whenever you see a bottleneck, it helps with bottleneck identification.”
Forward-looking simulation improves proactivity	Monitoring becomes more valuable when extended with simulation or forecasting to anticipate near-future bottlenecks.	“What I would improve is taking it one step further by simulating what is coming, rather than just monitoring current performance.”
Early warning extends reaction time	Predicting disruptions several days in advance increases the time available to react before disturbances escalate.	“If you can identify this two or three days in advance, the warnings give you much more time to react to the disturbance before it becomes an actual disruption.”
From information to action	Monitoring information creates value only when followed by downstream action-oriented mechanisms.	“Once you have the information, you have to take action.”
Action-executing agent extension	A natural extension of monitoring agents is to build agents that execute operational actions based on the detected signals.	“A new agent built on top of this one would start executing those actions.”

Continued on next page.

Table C.4 continued.

Code Name	Definition	Source Extract
Documentation control as risk mitigation	Documentation and reporting agents primarily reduce risk by checking mismatches and inconsistencies in internal records.	"I would score risk as a two and the rest as zero."
Data mismatch threatens finance and relationships	Poor documentation quality can create financial inaccuracies and damage client relationships.	"Mismatches in internal data sources or warehouse management systems can lead to inaccurate finances and reporting, which risks payments and client relations."
Document quality and standardization	Documentation agents support report quality, invoice accuracy, and document standardization.	"This agent mostly helps with report quality, invoice data quality, and document standardization."
Limited direct network resilience impact	Documentation agents may support administrative resilience without strongly affecting physical network resilience.	"From a finance and documentation perspective, yes, it helps a lot. However, on the network resilience side, there is not much exposure."
Master data as single source of truth	Documentation agents should be connected to centralized master data systems to improve consistency and learning.	"An important improvement would be connecting it to a centralized database, like Master Data Management, which serves as a single source of truth."
Feedback loop for continuous correction	Corrections made by humans should be fed back into the data system so repeated errors can be reduced over time.	"Whatever corrections those people make should feed back into the master data management system so the agent can learn."
Fixing errors at the source	Repeated data problems should be addressed at their origin rather than only corrected downstream.	"If an error is repetitive, I would try to fix it at the source."
Customer-side resilience through matching accuracy	Better matching between orders and documentation can improve resilience on the customer-facing side of operations.	"Ensuring the correct matching between orders and documentation helps improve resilience on the customer side."
Feedback reveals hidden supply chain issues	Customer feedback can uncover unknown supply chain problems such as delays or unmet service expectations.	"You can identify unknown issues in your supply chain, like product delays or unmet delivery preferences."
Blind spot identification through feedback	Feedback mechanisms help reveal operational blind spots that would otherwise remain invisible to the company.	"You can identify blind spots that you wouldn't find otherwise."

Continued on next page.

Table C.4 continued.

Code Name	Definition	Source Extract
Indirect contribution of feedback to resilience	Feedback agents contribute to resilience indirectly by improving visibility on issues rather than directly stabilizing network operations.	"I think this agent helps company resilience indirectly."
Separation of sensing and acting functions	Feedback collection and transactional actions should be split into distinct specialized agents.	"I would separate it into two specialized agents: one for capturing feedback to identify issues, and another for taking actions like processing reimbursements."
Disturbance-oriented expert support	Expert guidance agents are most useful once a disturbance or issue has emerged and someone needs structured support.	"I think it would be most useful for a disturbance, when you start seeing an issue."
On-demand explanation and knowledge capture	Guidance agents create value by summarizing expert knowledge and explaining what to do in uncommon situations.	"It helps most with giving explanations when something happens. Instead of calling an expert, one agent summarizes everything and builds knowledge for specific issues."
Monitoring-triggered guidance	Expert guidance becomes more valuable when connected to monitoring systems that automatically push recommendations.	"I would connect it to a monitoring agent. When it sees specific issues, it could automatically send suggestions to the relevant owner."
Demand-supply mismatch management	Forecast support agents help address mismatches between expected demand and operational supply capability.	"The problem it helps with most is the mismatch between demand and what you can offer."
Risk-oriented forecasting support	Demand forecast support is mainly proactive and therefore strongly linked to risk mitigation.	"I would give a higher score to the risk side—a two—and a one to disturbance."
Scenario-based planning need	Demand planning should not rely on a single forecast but incorporate multiple future scenarios.	"Another improvement is scenario definition; you often need multiple scenarios for the future, not just one forecast."
Marketing as resilience response lever	Demand-support agents can be linked to marketing actions to respond when demand is below expectations.	"If there is a mismatch or lower demand, the marketing agent could start promotional actions to increase demand."
Dashboard consolidation saves time	Insights agents create value by combining many data sources and reducing the time needed to interpret operations.	"It helps identify blind spots in operations and saves a lot of time by combining different data sources to generate insights faster."

Continued on next page.

Table C.4 continued.

Code Name	Definition	Source Extract
Insights as basis for action items	Insight generation should be extended into explicit action creation rather than stopping at analysis.	"I would build an agent that creates action items on top of these insights."
Inventory and waste risk visibility	Forecast and waste monitoring agents help identify risks related to low inventory, high demand, and upcoming product expiration.	"When you see high demand and low inventory, you know it can cause disruptions... you can identify upcoming expirations."
Supply and inventory level management	Waste monitoring and forecast monitoring support resilience by improving supply and inventory management.	"It mostly helps with managing supply and inventory levels."
Combining expiration with demand curves	Perishability monitoring should be integrated with demand patterns to make inventory decisions more proactive.	"I would make it more proactive by combining expiration dates with upcoming demand."
Dynamic ordering adjustment	Demand changes should inform whether to reduce or stop replenishment of newer products.	"By following the demand curve, you can adjust inventory levels and stop ordering newer products if demand drops."
Claims management across disturbance and disruption	Claims-related agents help manage problems that span from emerging disturbances to realized disruptions.	"I would give zero for uncertainty, one for risk, two for disturbance, and one for disruption."
Cash flow gap reduction	Claims agents reduce mismatches between expected and actual cash flows.	"It eliminates the gap between actual cash flow and expected payments."
Payment behavior for expectation setting	Historical payment behavior can be used to build more anticipatory agents for financial uncertainty.	"I would focus on the uncertainty aspect by building an agent based on historical payment behavior to set expectations."
Broad AI application areas for resilience	AI opportunities for resilience span network design, balancing, reallocation, customer service, insights, and data quality.	"Network design, supply-demand balancing, resource reallocation, customer service, insights, and data quality."
Quality checks as safeguard	Quality-control mechanisms are essential safeguards for AI deployment in resilience-related processes.	"It is crucial to have quality checks and a human-in-the-loop."
Human-in-the-loop for hallucination control	Human oversight remains necessary because GenAI can produce outputs that appear convincing but are factually wrong.	"GenAI can produce very realistic but inaccurate results."

Continued on next page.

Table C.4 continued.

Code Name	Definition	Source Extract
Process harmonization across teams	AI guidelines must be aligned with existing processes and consistently coordinated across organisational teams.	"The guidelines should be harmonized with your processes, ensuring alignment across all teams."
Formal sign-off for process change	Changes to resilience-related processes should require explicit sign-off from all relevant stakeholders.	"One guideline should involve getting sign-off from everyone to formally change processes within the resilience team."

Themes and Supporting Open Coding Names

Table C.5: Themes and Supporting Open Coding Names – Second Interview (Supply Chain Expert)

Theme Name	Associated Codes
Theme 1: Monitoring and Visibility as the Basis of Resilience Action	Interrelated resilience dimensions; Disturbance-oriented monitoring value; Historical monitoring supports risk detection; Predictive maintenance support; KPI-based operational tracking; Bottleneck identification through flow visibility; Dashboard consolidation saves time; Inventory and waste risk visibility; Supply and inventory level management
Theme 2: From Detection to Proactive and Action-Oriented Intervention	Forward-looking simulation improves proactivity; Early warning extends reaction time; From information to action; Action-executing agent extension; Monitoring-triggered guidance; Scenario-based planning need; Insights as basis for action items; Combining expiration with demand curves; Dynamic ordering adjustment; Marketing as resilience response lever
Theme 3: Data Quality, Standardization, and Learning Loops	Documentation control as risk mitigation; Data mismatch threatens finance and relationships; Document quality and standardization; Master data as single source of truth; Feedback loop for continuous correction; Fixing errors at the source; Customer-side resilience through matching accuracy; Broad AI application areas for resilience
Theme 4: AI Support for Disturbances, Mismatches, and Operational Blind Spots	Feedback reveals hidden supply chain issues; Blind spot identification through feedback; Indirect contribution of feedback to resilience; Disturbance-oriented expert support; On-demand explanation and knowledge capture; Demand-supply mismatch management; Risk-oriented forecasting support; Claims management across disturbance and disruption; Cash flow gap reduction; Payment behavior for expectation setting
Theme 5: Functional Specialization and Connected Agent Architectures	Separation of sensing and acting functions; Action-executing agent extension; Monitoring-triggered guidance; Marketing as resilience response lever; Insights as basis for action items; Payment behavior for expectation setting
Theme 6: Governance, Human Oversight, and Organisational Alignment	Anonymity and non-client framing; AI as support for summarization; Limited direct network resilience impact; Quality checks as safeguard; Human-in-the-loop for hallucination control; Process harmonization across teams; Formal sign-off for process change

Thematic Analysis

Table C.6: Themes and Definitions – Second Interview (Supply Chain Expert)

Theme Name	Definition
Theme 1: Monitoring and Visibility as the Basis of Resilience Action	This theme captures the expert's repeated emphasis on monitoring, visibility, and information consolidation as the core starting point for resilience. Across capacity, insights, forecast, and waste-related agents, the value of AI lies first in making problems visible: identifying exceeded thresholds, detecting trends in historical data, highlighting bottlenecks, revealing blind spots, and clarifying supply and inventory risks. The expert consistently frames resilience support as beginning with the ability to observe the system through KPIs, dashboards, stock signals, and operational flows. In this sense, AI is valuable not only because it automates analysis, but because it transforms fragmented data into actionable visibility on disturbances and emerging risks.
Theme 2: From Detection to Proactive and Action-Oriented Intervention	This theme reflects the expert's view that monitoring alone is insufficient unless it is extended into anticipation and action. Several times, the expert argues that agents should go beyond observing current conditions and instead simulate what is coming, issue warnings earlier, and trigger concrete responses. This includes forecasting bottlenecks days in advance, generating action items from insights, combining demand with expiration data, and linking forecasting outputs to commercial or operational levers such as marketing or replenishment adjustment. The central logic is that resilience is improved when AI shifts organisations from reactive visibility to proactive intervention, giving them more time and more structured ways to respond before disturbances become disruptions.
Theme 3: Data Quality, Standardization, and Learning Loops	This theme concerns the role of data integrity as a foundational enabler of resilience-oriented AI. In the expert's view, documentation agents are especially valuable for reducing risks related to data mismatches, inaccurate invoices, inconsistent reporting, and poor standardization. However, the expert also argues that these systems should not function only as downstream checking mechanisms. Instead, they should be connected to centralized master data structures, incorporate human corrections through feedback loops, and ultimately fix recurring issues at their source. The theme therefore frames resilience not only as reacting to operational shocks, but also as building reliable informational infrastructure that improves decision quality, reduces recurring administrative errors, and strengthens customer-facing consistency.
Theme 4: AI Support for Disturbances, Mismatches, and Operational Blind Spots	This theme captures the expert's interpretation of many AI applications as tools for handling mismatches and making hidden operational issues visible. Feedback agents reveal unknown service problems and unmet customer expectations; guidance agents provide knowledge when unusual issues occur; forecasting agents address mismatches between demand and what the system can offer; and claims agents reduce gaps between expected and actual payments. Across these examples, AI is not primarily framed as eliminating uncertainty altogether, but as helping organisations identify problems once they begin to emerge and respond more effectively. The theme therefore positions AI as a mechanism for surfacing blind spots, clarifying disturbances, and supporting response when operational reality deviates from expectations.

Continued on next page.

Table C.6 continued.

Theme Name	Definition
Theme 5: Functional Specialization and Connected Agent Architectures	This theme reflects the expert's preference for architectures composed of specialized yet interconnected agents rather than overly broad, monolithic systems. The expert explicitly recommends separating feedback capture from reimbursement execution, connecting guidance to monitoring, adding action-execution layers on top of monitoring, linking demand forecasting to marketing responses, and building expectation-setting agents from payment histories. This suggests a modular understanding of AI design, where different agents serve distinct roles such as sensing, interpreting, recommending, or acting, but generate greater value when these roles are connected. Resilience, from this perspective, is supported not by a single all-purpose agent, but by an ecosystem of coordinated functions aligned to specific tasks.
Theme 6: Governance, Human Oversight, and Organisational Alignment	This theme captures the governance-oriented conditions the expert sees as necessary for responsible AI deployment. At the methodological level, the expert stresses anonymity and non-client framing, reinforcing the importance of confidentiality and professional boundaries. At the operational level, the expert emphasizes quality checks and human-in-the-loop safeguards because GenAI can produce outputs that appear convincing while still being inaccurate. Finally, at the organisational level, the expert argues that guidelines must be harmonized with business processes, aligned across teams, and formally approved through sign-off when resilience-related processes are changed. Altogether, the theme frames effective AI adoption not as a purely technical matter, but as one that depends on oversight, process integration, and structured organisational governance.

C.3. Interview 1 - AI Expert

Interview Summary

The interviewee differentiates GenAI from Agentic AI based on task complexity. GenAI is appropriate for simple, low-dependency tasks such as generating summaries or emails, where errors can be readily identified. In contrast, Agentic AI is designed for multi-step workflows that demand planning, verification, and explicit monitoring. The deployment of agentic systems necessitates foundational readiness, including high-quality data, robust system integrations, continuous monitoring, and dedicated testing environments. The interviewee characterizes regulations as contextual constraints rather than fundamental technical requirements. Regarding safeguards, the interviewee advocates for human supervision instead of continuous human-in-the-loop involvement, thereby maintaining automation benefits through periodic oversight rather than stepwise approval. Key safeguards identified include clear accountability frameworks, comprehensive logging, and well-defined fallback procedures. The interviewee notes that explainability requirements depend on stakeholder needs but should generally clarify why a particular outcome occurred and why alternatives were not chosen. For agentic systems, explanations are necessary at both individual decision points and the overall system level. Concerning sustainability, the interviewee provides limited commentary, observing only that GenAI may have a smaller environmental footprint compared to continuously operating agentic systems. In terms of method selection, the interviewee recommends standard machine learning for prediction and ranking tasks, GenAI for summarization and document matching, and Agentic AI exclusively for genuinely complex multi-step reasoning. However, the interviewee observes that current agentic systems are still immature and require substantial human involvement. Multi-layered architectures, in particular, necessitate robust safeguards to prevent system failures.

Open Coding

Table C.7: Open Coding Analysis – First Interview (AI Expert)

Code Name	Definition	Source Extract
Generative artificial intelligence suitability	Generative artificial intelligence is best suited for bounded tasks with limited dependencies, where outputs can be quickly reviewed and corrected by humans.	“tasks that are relatively small and do not require many dependencies or complex connections. Typical examples are analysing information, producing summaries, or drafting emails. These are single-step or short multi-step tasks where any errors can be easily identified and corrected by a human.”
Short workflows	Generative artificial intelligence fits workflows that are either single-step or only require a short sequence of actions, without complex coordination.	“These are single-step or short multi-step tasks where any errors can be easily identified and corrected by a human.”
Easy human correction	When mistakes can be detected and fixed easily by a person, lightweight generative artificial intelligence support is typically sufficient.	“any errors can be easily identified and corrected by a human.”
Plan act and check loop	Agentic systems resemble a worker that must plan, take actions, and verify outcomes, rather than producing a single output.	“You can think of it as a real person performing a complicated task: it needs to plan, to act, and to check different things”
Monitoring layer	Agentic artificial intelligence should operate under a monitoring or oversight layer that supervises behaviour and constrains failures.	“while operating under some monitoring layer that oversees what it does.”
Reliable data requirement	Agentic deployments require dependable, high-quality data; without it, meaningful automation is not feasible.	“The first and most important requirement is reliable data; you need good-quality data to build anything meaningful.”
Robust system interfaces	Agents require stable system interfaces to connect components without fragile, break-prone integrations.	“The system should have robust interfaces that allow the agent to connect to different components without breaking easily.”
Testing environment	Before release, agents should be validated in a testing environment to verify safe deployment, monitoring effectiveness, and performance.	“it is also important to have a testing system or testing environment where you can verify that the agents you build can be deployed safely, monitored effectively, and that they actually perform the intended job properly.”

Continued on next page.

Table C.7 continued.

Code Name	Definition	Source Extract
Continuous monitoring	Agentic systems need ongoing monitoring to detect drift, failures, or unexpected actions during operation.	"there should be continuous monitoring of its behaviour."
Regulatory context	Regulatory aspects affect deployment and accountability, but are not framed as part of the agent's core technical mechanism.	"Regulations and similar aspects can also play a role, but they are less about the technical core of the Agentic AI itself."
Human in the loop approval	Human in the loop means a human must verify or approve each decision before it is applied, directly shaping system actions.	"the agent or model produces a decision or solution, and a human must verify or approve it every time before it is applied."
Human in the loop participation	In human in the loop setups, humans are part of task execution rather than external overseers.	"the human is actively part of the task execution."
Human supervision	Human supervision means monitoring performance and intervening when needed, without reviewing every output.	"The human monitors the system periodically, checks for visible errors or issues, and intervenes when something needs to be corrected, but is not involved in every single decision."
Supervision preferred	For agentic systems, supervision is preferred because constant approvals reduce the value of automation.	"I would suggest human supervision, because the goal of an Agentic AI setup is usually to automate work. If you keep a human in the loop on every decision, you reduce the benefits of that automation."
Accountability framework	A foundational safeguard is explicit accountability: defining who owns outcomes and responsibility when failures occur.	"You need to know where responsibility lies if something goes wrong, and which person or organisation is accountable."
Monitoring and logging	Operational acceptability depends on recording actions so incidents can be reconstructed and audited.	"proper monitoring and logging are crucial. The system's actions should be recorded so that you can trace what happened."
Fallback options	Agents should have predefined fallback options to handle failures or non-performance without operational breakdown.	"There should also be defined fallback options or contingency plans: what happens if something goes wrong, and what happens if the agent fails to perform a task. Having these 'Plan B' paths in place is an important part of making autonomous operation acceptable."
Stakeholder dependent explainability	Useful explanations vary by role; different stakeholders need different forms of reasoning and evidence.	"The type of explanation that is useful depends strongly on the stakeholder. Different roles require different information."

Continued on next page.

Table C.7 continued.

Code Name	Definition	Source Extract
Planning role explanations	Planning stakeholders focus on what factors drive outcomes and why a specific prediction was produced.	“Someone working on demand forecasting or planning might care about how the relevant features relate to the outcome and why the system produced a specific prediction.”
Data scientist explanations	Technical stakeholders may prefer performance metrics and model-level behaviour to assess reliability.	“A data scientist, on the other hand, might focus more on model-level aspects such as performance metrics and internal behaviour.”
Contrastive explanations	A common explainability need is understanding why the chosen outcome occurred and why alternatives did not.	“stakeholders generally want to understand why a particular outcome is happening and why alternative outcomes are not happening.”
Local level explainability	Each individual decision should be explainable at the point it is made, supporting accountability and review.	“At the local level, you should be able to explain why each individual decision was made at each decision point.”
System level explainability	Beyond single decisions, stakeholders need clarity on overall behaviour, component interactions, and decision chains.	“At the system level, you should also be able to explain how the overall system behaves, how the different components interact, and how the chain of decisions leads to a given outcome.”
Limited sustainability expertise	The interviewee frames sustainability comments as intuitive rather than expert assessment.	“Sustainability is not really my area of expertise, so I can only give an intuitive view.”
Generative artificial intelligence compute profile	Generative artificial intelligence is perceived as more contained when training is amortised and usage is mostly inference.	“If you train a GenAI model once and then mainly use it for inference, it feels like the total computational load can be relatively contained.”
Agentic artificial intelligence compute profile	Agentic systems may demand more compute due to continuous decision-making and ongoing operation.	“Agentic AI, by contrast, often runs continuously, making decisions and acting over time.”
Higher environmental impact risk	The interviewee suggests that continuous agent operation could imply higher environmental impact, while explicitly noting uncertainty.	“From common sense, that suggests it needs more computational power and therefore may have a higher environmental impact, but I would not claim detailed expertise on the sustainability side.”
No sustainable by design guidance	The interviewee could not propose specific design guidance for sustainable by design systems within the agent context.	“When asked about designing a ‘sustainable-by-design’ agent, I could not really provide an answer because that topic is outside my domain.”

Continued on next page.

Table C.7 continued.

Code Name	Definition	Source Extract
Speed on repetitive tasks	Artificial intelligence tends to outperform humans when tasks are clearly defined, repetitive, and can be executed at scale without fatigue.	"AI tends to perform very well when it is assigned a single well-defined task. ... for repetitive tasks, computer systems have an advantage because they do not get tired and can operate at large scale."
Pattern detection at scale	Artificial intelligence can identify patterns in very large datasets more effectively than humans, especially under scale.	"It is also particularly strong at detecting patterns in very large datasets, something humans are not good at."
Large scale summarisation	Once configured, automated systems can summarise large volumes quickly without losing quality.	"if I had to summarise a thousand documents, it might take me a very long time, whereas a computer can do it extremely quickly without a loss in quality once the system is properly set up."
Speed exploration trade off	Excessive speed in exploration can yield fragile or suboptimal outcomes due to the exploration and exploitation balance.	"If you 'explore' too aggressively or move too fast through the search space, you may end up stuck in a wrong local minimum or maximum."
Machine learning for ranking	When outputs are simple predictions used for ranking or prioritisation, standard machine learning is described as sufficient.	"When you are talking about prediction, where the prediction is essentially a simple output that you can use to rank or prioritise things, then I would consider that to be standard machine learning."
Rules on predictions	Lightweight rule layers can sit on top of machine learning predictions without requiring agentic reasoning.	"You might put some simple rules on top of the predictions –for example, 'the first one goes, then the second one' –but there is nothing that requires more than that."
Generative artificial intelligence for matching	Generative artificial intelligence is positioned as suitable for summarisation and straightforward document matching tasks.	"summarising information or doing straightforward matching of documents against some desired outcome, that, for me, falls under GenAI."
Agentic artificial intelligence for complexity	Agentic artificial intelligence becomes relevant when tasks involve many steps, multiple levels, and substantial reasoning.	"Agentic AI becomes relevant when there is real complexity: different levels, many steps, and a lot of reasoning involved."
Simpler methods sufficient	Even if agentic systems could perform prediction or summarisation, the interviewee argues simpler methods frequently suffice.	"That does not mean that Agentic AI *cannot* also handle prediction or summarisation, but in many of these cases simpler methods are sufficient."

Continued on next page.

Table C.7 continued.

Code Name	Definition	Source Extract
Technology not mature	The interviewee expresses caution about current maturity, anticipating continued human involvement for many use cases.	"Moreover, I am not sure that our current agentic AI technology is fully there yet. For many of the use cases, I would still expect a significant amount of human involvement."
Organisational analogy	Multi-layer agentic systems are framed as analogous to organisational hierarchies with distributed control and responsibility.	"What does 'multi-layer' mean for a human organisation? It means you have different levels of control and responsibility. The same idea applies here"
Safeguards and oversight	Because multi-layer systems can fail in complex ways, strong safeguards and human oversight are viewed as necessary.	"A multi-layer agentic system requires a lot of reasoning ability and many safeguards to make sure that things do not fall apart."
Generative artificial intelligence failures	Observed failures even in lower-level generative artificial intelligence are used as an argument for cautious deployment of more complex agentic systems.	"Even with GenAI –which is a much 'lower-level' capability compared to full agentic systems –we already see many failures. So, while multi-layer agents are an appealing idea and mirror how humans organise themselves, in practice I would still be very cautious and would expect a strong need for human oversight."

Themes and Supporting Open Coding Names

Table C.8: Themes and Supporting Open Coding Names – First Interview (AI Expert)

Theme Name	Associated Codes
Theme 1: Task Complexity as Technology Selector	Generative artificial intelligence suitability; Short workflows; Easy human correction; Machine learning for ranking; Rules on predictions; Generative artificial intelligence for matching; Agentic artificial intelligence for complexity; Simpler methods sufficient
Theme 2: Operational Infrastructure Requirements for Agentic AI	Reliable data requirement; Robust system interfaces; Testing environment
Theme 3: Human Oversight Configurations	Human in the loop approval; Human in the loop participation; Human supervision; Supervision preferred
Theme 4: Accountability and Safeguard Mechanisms for Agentic AI	Accountability framework; Monitoring and logging; Fallback options; Continuous monitoring; Safeguards and oversight; Regulatory context
Theme 5: Multi-Stakeholder Explainability	Stakeholder dependent explainability; Planning role explanations; Data scientist explanations; Contrastive explanations; Local level explainability; System level explainability

Continued on next page.

Table C.8 continued.

Theme Name	Associated Codes
Theme 6: Agentic AI System Architecture	Plan act and check loop; Monitoring layer; Organisational analogy; Technology not mature; Generative artificial intelligence failures
Theme 7: AI Performance Advantages	Speed on repetitive tasks; Pattern detection at scale; Large scale summarisation; Speed exploration trade off

In this section, the thematic analysis of the two AI interviews is presented, followed by a section that compares the current findings with the previous ones.

Thematic Analysis

Table C.9: Themes and Definitions – First Interview (AI Expert)

Theme Name	Definition
Theme 1: Task Complexity as Technology Selector	This theme examines how task nature and complexity determine the selection of appropriate paradigm. Standard machine learning is sufficient for tasks involving simple predictions used for ranking or prioritization, which may be enhanced with lightweight rule-based layers. Generative AI is most effective for bounded, single-step, or short multi-step tasks with limited dependencies, such as summarization, document matching, and drafting communications, where errors can be readily identified and corrected by humans. Agentic AI is warranted only for tasks characterized by significant complexity, involving multiple levels, numerous steps, and substantial reasoning. The guiding principle is to prefer simpler methods when adequate, reserving Agentic AI for situations where complexity necessitates its use.
Theme 2: Operational Infrastructure Requirements for Agentic AI	This theme addresses the essential technical and organisational prerequisites for deploying Agentic AI systems. Reliable, high-quality data is the primary requirement, as meaningful automation cannot be achieved without it. Robust system interfaces are necessary to connect components and prevent fragile integrations. A dedicated testing environment is required to verify that agents can be deployed safely, monitored effectively, and perform their intended functions. Collectively, these infrastructure elements establish the technical foundation necessary for successful Agentic AI deployments.
Theme 3: Human Oversight Configurations	This theme differentiates modes of human involvement in AI systems and evaluates their suitability for specific contexts. Human-in-the-loop configurations require verification or approval of each decision prior to implementation, making the human an active participant in task execution. In contrast, human supervision entails periodic monitoring, detection of observable errors, and intervention only when necessary. For Agentic AI systems designed to automate tasks, supervision is generally preferred, as constant approvals would reduce the advantages of automation. In Generative AI applications, human-in-the-loop arrangements may be more suitable because errors can be readily identified and corrected. The selection between these approaches represents a fundamental trade-off between control and efficiency.

Continued on next page.

Table C.9 continued.

Theme Name	Definition
Theme 4: Accountability and Safeguard Mechanisms for Agentic AI	This theme examines the governance structures and protective measures required for the responsible autonomous operation of Agentic AI. Explicit accountability should delineate ownership of outcomes and assign responsibility in the event of failures. Comprehensive monitoring and logging facilitate tracing system actions for incident reconstruction and auditing. Predefined fallback options and contingency plans support operational continuity when agents fail or encounter unforeseen circumstances. In multi-layer Agentic AI systems, where failures may occur in complex ways, these safeguards are particularly essential. This theme adopts a cautious perspective, emphasizing robust protective measures in light of documented failures in even simpler Generative AI applications.
Theme 5: Multi-Stakeholder Explainability	This theme explores how diverse audiences require tailored forms of explanation from AI systems, including those based on Machine Learning, Generative AI, and Agentic AI. The nature of useful explanations depends on the audience: planning stakeholders emphasize feature-outcome relationships and prediction rationale, whereas data scientists prioritize performance metrics and model-level behavior. A consistent requirement across roles is contrastive explanation, which involves clarifying why a specific outcome occurred instead of alternatives. Explainability should operate at both local and system levels. Individual decisions must be explainable at each decision point, and overall system behavior, component interactions, and decision chains should be transparent. In the context of Agentic AI, system-level explainability is especially critical due to the complexity of multi-step reasoning and component interactions.
Theme 6: Agentic AI System Architecture	This theme characterises the distinctive operational nature of Agentic AI that differentiates it from Generative AI and Machine Learning. Agentic AI systems function through a plan-act-check loop, analogous to a real person performing complicated tasks: planning actions, executing them, and verifying outcomes. Unlike Generative AI, which produces single outputs, Agentic AI systems operate continuously under a monitoring layer that oversees behaviour and prevents failures. Multi-layer Agentic AI systems mirror organisational hierarchies with distributed control and responsibility. This architectural complexity demands substantial reasoning ability and robust safeguards, with the technology acknowledged as not yet fully mature for many use cases. Even Generative AI—a simpler capability than full Agentic AI systems — already exhibits many failures, warranting caution for more complex deployments.

Continued on next page.

Table C.9 continued.

Theme Name	Definition
Theme 7: AI Performance Advantages	This theme delineates the specific conditions in which artificial intelligence (AI) surpasses human performance, a phenomenon observed across various AI technologies. AI demonstrates superior proficiency in executing clearly defined, repetitive tasks at scale without experiencing fatigue, a capability relevant to Machine Learning, GenAI, and Agentic AI. The identification of patterns within extensive datasets exemplifies an area where human limitations are significant, and Machine Learning approaches are particularly effective. Additionally, large-scale summarisation, enabled by GenAI, illustrates how well-configured systems can efficiently process substantial volumes of information without compromising quality. Nevertheless, excessive speed in exploration may result in suboptimal outcomes due to the exploration-exploitation trade-off, indicating that processing speed must be carefully balanced with solution quality in any AI application.

C.4. Interview 2 - AI Expert

Interview Summary

The interviewee characterizes the distinction between GenAI and Agentic AI as primarily operational rather than technological, given that most Agentic AIsystems are built upon generative models. Generative configurations are sufficient for producing information or suggestions, whereas Agentic AIconfigurations are suitable when the system is required to function as a trained teammate, executing tasks end-to-end within defined constraints. Successful deployment is determined more by organizational preconditions than by model capability alone. While robust models are necessary, privacy and responsible use are foundational requirements. Significant value is realized when these tools are integrated with internal backbone systems such as ERP platforms, allowing retrieval-augmented generation to utilize relevant internal knowledge instead of producing generic outputs. Human involvement should serve as a control mechanism within the system architecture. Human-in-the-loop processes are primarily responsible for validation and exception handling, particularly during early deployment phases when outputs require frequent review. Although the need for validation may diminish as systems mature, it should not be eliminated entirely, as models necessitate ongoing maintenance and periodic re-validation. Early workforce involvement is essential for building organizational trust, as users possess domain expertise and must develop confidence prior to relying on the system. Autonomy should be governed by risk-based thresholds: routine, low-impact actions may be automated, whereas high-impact decisions necessitate escalation. Rapid decision-making introduces risk in exceptional cases with significant exposure; therefore, structured exception flows and clearly defined boundaries are essential. Explainability is essential for both safety and adoption. Stakeholders require clear reasoning behind decisions and transparency regarding the factors considered or omitted. In multi-agent chains, visual explanations are preferred; however, providing evidence for the system's decision-making is critical. Explainability also facilitates continuous improvement through human stress-testing and feedback mechanisms. Regarding disruption management, the interviewee advises against relying primarily on historical examples, as disruptions are multi-dimensional and may not be represented in past data. Organizations should explicitly model relevant disruptions, the mechanisms for sensing them, and the governance of response strategies. Simulation is necessary for exploring scenarios beyond historical precedent, but it must be informed by domain expertise and validated by multiple stakeholders. A critical capability is the recognition of when the system operates beyond its established parameters. Low confidence levels should prompt human intervention, while high confidence may justify automation. Escalation functions as both a safety measure and a learning opportunity, with corrections feeding back into the system. Robustness evaluation requires assessment of the entire process, from disruption sensing to decision execution and governance, with emphasis on enhanced sensing, comprehensive scenario exploration, and user

interfaces that facilitate decision transparency.

Open Coding

Table C.10: Open Coding Analysis – Second Interview (AI Expert)

Code Name	Definition	Source Extract
GenAI–agent link	GenAI and agents are framed as tightly connected, with agents built on generative models.	"I think they are nowadays quite, very much related to each other …agents built on top of the current system."
Task-trained agents	Agents are valued when they can be trained for clear, repeat tasks.	"The benefits of having agents that you can really train them in specific tasks …"
Model access	Strong language models are seen as a basic need before deployment.	"Yeah, of course. The access to good, good …models. I think that's important."
Privacy setup	Safe use needs privacy controls and secure handling of company data.	"There is the the privacy …we have own host version …so you can safely upload data."
Use it right	Responsible use is needed to get real value from the tool.	"Responsible uses of it and to really get all the benefits of it …"
Link to systems	Value rises when AI is linked to internal tools, not left standalone.	"I like to have connected with all the older internal systems."
Core system link	Connecting to core systems is described as what unlocks the full potential.	"When you start connecting it to all different kinds of backbone systems …for example doing the invoicing, then I think you really unlock the potential."
RAG option	Retrieval is accepted as a practical way to use internal info.	"For example using like the rag, I could be an option." "Yeah. For example, yeah."
Human control	Humans are needed as control and for handling exceptions.	"I think really as a control system or basically control and exception."
Early checks	At the start, humans should validate outputs often.	"In the start where you deploy an agent, you need to do that quite often."
Fewer checks later	As the system matures, checks can be reduced.	"The moment that gets more mature of time that you might can reduce the validation."
Keep maintaining	Agents must be maintained and re-checked over time.	"You cannot simply just deploy it and stay there forever. I think you have to maintain it, validate it and confirm it."
Build trust	Trust in the tool must grow in the workforce.	"There's also the trust that needs to grow into the system."
Use expert input	Users should be involved because they hold the know-how for real cases.	"They need to use it …they also have the knowledge for how to react …because they have the expertise."

Continued on next page.

Table C.10 continued.

Code Name	Definition	Source Extract
Risk thresholds	Autonomy should depend on impact; high impact needs review.	"Let's say the exceptions. That's only a cost of 1000. Then I'm fine with it …The cost impact is more than a million …you want some second pair of eyes."
Guardrails	Clear limits and exception flows should be designed in advance.	"I think you should build in some boundaries or guardrails and then get into exception flows."
Show the why	Stakeholders need the reason behind actions and non-actions.	"At least the reasoning …the human should understand? Why search decision has been made or have not been made?"
Say what was missed	Good explainability also covers what was not used in the decision.	"Why are those no take into accounts?"
Stress test	Humans should stress-test key decisions.	"It's up to the human to also do so …stress testing of the decision being made …"
Feed back	New info from review should be added back to improve the system.	"You should see that the new information into your algorithm because then you're strengthening your algorithm of time …"
Show evidence	Explainability should include evidence; format can vary.	"I like to make it visual. Can we also be text? …As long as the evidence is provided …providing the evidence that that's key."
Repeat tasks	Speed gains fit repeatable actions.	"Repeatable actions, I would say …"
Data heavy	Agents are faster when lots of info must be processed.	"Actions were a lot of data as needed."
Many agents	Multiple agents can cover different areas and be combined.	"With multiple agents …you can really train them in a very specific area …combining them, I think that's something really powerful."
Speed risk	Speed becomes risky in high-stakes exception cases.	"Again, in those exception cases where a lot of money is on the table."
Disruption is wide	Disruption is framed as multi-layer and not one single case.	"Disruption is not an one thing …disruption has multiple layers and multiple dimensions."
Model disruption	Disruptions must be defined and sensed, not assumed in history.	"You need to start modelling what is the disruption …you need to start sensing it in your data."
Sim is key	Simulation is presented as key for training and tuning.	"I think simulation is the key part to train a lot to get the parameters right …"

Continued on next page.

Table C.10 continued.

Code Name	Definition	Source Extract
Make missing cases	Synthetic data helps cover cases not found in historical data.	"You can make up your own …1 disruption that you want to support …if you just look to historical data …you might not gonna find them."
Human guides	Humans must define what to sense and how to act in disruption.	"Beginning human intelligence to really train them all. OK, what is the description? How to sense that how to act upon it."
Many validators	Validation should involve multiple people to build trust and quality.	"It should also do with multiple people, and again that is also for building trust into the system."
Use confidence	Confidence levels should drive handover vs automation.	"If you see that the ADIAI model only has 10% confidence …hand it over to the planner …However, then 99% …you might start automating …"
Try then escalate	When unsure, the system should try to solve, but still involve humans.	"I think it's gonna be both. Trying to solve it …but in the end we have the human in the loop …"
Learn from wrong	Wrong calls should feed back into the system for next time.	"If the decision was not right, let's feedback the new information …from the human back into the system …next time it might be first time, right?"
Full-cycle test	Robustness is judged across sensing, options, and action as a full loop.	"It's not finding that unexpected event, it's for the full cycle …how to sense it …to determine the right options, there's some scenario play."
Top focus areas	Key priorities are sensing, options, and scenario work.	"That really comes back to sensing, sensing options and scenario play."
Make it clear	The system should support humans and make decisions easy to grasp.	"Make that interface with the human …not only make decisions, but also make decisions understandable."

Themes and Supporting Open Coding Names

Table C.11: Themes and Supporting Open Coding Names – Second Interview (AI Expert)

Theme Name	Supporting Open Coding Name
Theme 1: GenAI and agents as one stack	GenAI–agent link; Task-trained agents; Model access
Theme 2: Enterprise readiness as privacy plus system links	Privacy setup; Use it right; Link to systems; Core system link; RAG option
Theme 3: Human oversight as a control loop	Human control; Early checks; Fewer checks later; Keep maintaining; Build trust; Use expert input

Continued on next page.

Table C.11 continued.

Theme Name	Supporting Open Coding Name
Theme 4: Risk-based autonomy and guardrails	Risk thresholds; Guardrails; Speed risk
Theme 5: Explainability requires reasons and evidence	Show the why; Say what was missed; Show evidence; Stress test; Feed back
Theme 6: Speed gains are real but bounded	Repeat tasks; Data heavy; Many agents; Speed risk
Theme 7: Robustness from modelling, sim, and confidence	Disruption is wide; Model disruption; Sim is key; Make missing cases; Human guides; Use confidence; Try then escalate; Learn from wrong; Full-cycle test; Top focus areas; Make it clear

Thematic Analysis

Table C.12: Themes and Definitions – Second Interview (AI Expert)

Theme Name	Definition
Theme 1: GenAI and agents as one stack	The participant treats Agentic AI as the operational layer built on top of GenAI, where the decision is not about labels but about whether the system must only generate content or also execute task steps reliably. In this view, agents matter when they can be configured like specialised workers that handle defined duties end-to-end, rather than producing one-off answers.
Theme 2: Enterprise readiness as privacy plus system links	Deployment readiness is conceptualized as the integration of secure data management and robust system connectivity. According to the participant, privacy-preserving configurations are essential for the safe utilization of internal data. However, significant business value is realized only when the model is integrated with internal tools and core platforms such as enterprise resource planning (ERP) systems. Once these connections are established, retrieval-based access is identified as an effective method to anchor outputs in internal data sources.
Theme 3: Human oversight as a control loop	Human involvement functions as a continuous control mechanism rather than serving solely as a temporary safety measure. According to the participant, early system deployments necessitate frequent validation to establish confidence in output quality. Although oversight may be reduced as the system matures, it remains essential because models require ongoing maintenance, re-validation, and monitoring. Trust-building is integral to this process, and domain experts are required both to evaluate outcomes and to guide appropriate system behavior.
Theme 4: Risk-based autonomy and guardrails	Autonomy is defined as contingent upon the level of impact. Explicit thresholds are established to distinguish low-impact actions, which proceed without escalation, from high-impact decisions that require review. This approach results in the implementation of guardrails and structured exception flows, directing high-risk cases to human oversight and ensuring that operational speed does not compromise accountability as the stakes rise.

Continued on next page.

Table C.12 continued.

Theme Name	Definition
Theme 5: Explainability requires reasons and evidence	Explainability is defined in practical terms, requiring that stakeholders understand the rationale behind actions taken or not taken, the information excluded, and the evidence supporting decisions. Additionally, explainability is connected to learning, as human stress-testing is expected to generate feedback that is incorporated into the system to enhance decision logic over time.
Theme 6: Speed gains are real but bounded	The participant identifies speed as a significant advantage for repetitive tasks and scenarios requiring rapid processing of large data volumes, particularly in multi-agent configurations where specialists address distinct domains. Conversely, speed is considered disadvantageous in exceptional high-stakes situations, as rapid automation may increase potential losses. These insights underscore the importance of integrating speed with risk controls and escalation procedures.
Theme 7: Robustness from modelling, sim, and confidence	Robustness is conceptualized not as prior exposure to disruptions, but as the implementation of an end-to-end process for defining disruptions, detecting them, evaluating response options, and determining escalation points. The participant highlights the use of simulation and synthetic scenarios to address gaps in historical data, and recommends confidence signals as a practical mechanism for identifying out-of-scope conditions and initiating human review. Instances in which humans override automated decisions should be incorporated into the system to enhance future performance, thereby establishing a continuous learning loop grounded in operational practice.

C.5. Final Agents Portfolio

Table C.13: Final portfolio of agents

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation (summary)
Capacity-risk assessment Agent	Continuously monitors utilisation/stress indicators across production and logistics nodes; produces short-term (day+1/day+2) capacity projections and risk scores; highlights emerging bottlenecks and escalates on threshold breaches.	Retrieve machine/node info; access utilisation histories and maintenance KPIs; generate short-term projections/risk scores; flag volume build-ups and downstream bottlenecks.	Modify production/routing schedules; change maintenance plans; shut down machines or re-route flows autonomously.	*Rollback:* Production/logistics operator cross-checks outputs vs. sensor readings and KPI definitions; repeated issues trigger joint review with maintenance and data engineering to recalibrate thresholds/sensors/models. *Escalation:* Notify line supervisors/planners at early-warning thresholds; escalate to production/logistics management when overload persists or critical resources are affected; escalate to IT/AI governance for systematic model errors.
Documentation / reporting Agent	Compiles consistent supply chain reports; gathers inputs from live systems and communications; reconciles against central MDM (single source of truth); flags significant mismatches and missing fields.	Reformat documents; reconcile content vs. WMS/DMS/ERP and MDM; flag inconsistencies/missing data; propose standardised templates.	Auto-send reports externally; overwrite master data; expose confidential information beyond access rights.	*Rollback:* Human reviewer (process owner/-data steward) validates, redacts/corrects, and approves final version; recurring data-quality issues are fed back into MDM to address root causes. *Escalation:* Notify business owner and data steward on inconsistencies; escalate to compliance for confidentiality breaches; escalate to MDM/data governance owners for structural data-quality problems.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Feedback Agent	Collects feedback from customers and internal actors; clusters/classifies signals to detect recurring pain points; prepares structured follow-up cases; can schedule returns and draft reimbursement requests (no transactions).	Access invoices/-client info within role-based permissions; collect feedback across channels; cluster/classify; create structured cases for returns/reimbursements.	Execute financial transactions; access/modify banking info; independently approve/reject claims.	*Rollback:* Handover to customer support or resilience analyst when cases are complex/ambiguous/misclassified; humans adjust classification, priority, and resolution. *Escalation:* Notify customer support/process owner for high-impact/high-value cases or systemic patterns; provide a dedicated channel/button to escalate to a human.
Expert-process guidance Agent	Guides rare/complex expert processes via structured prompts/templates; can be triggered on request or by other agents' signals; logs usage for improvement; escalates when missing processes or failures occur.	Access confidential procedural info on demand or predefined event triggers; provide step-by-step guidance; log guidance usage.	Access confidential info without consent/outside triggers; autonomously change core configuration parameters (e.g., pricing rules, regulatory flags).	*Rollback:* Control agent supervises queried sources; can halt workflow, mask sensitive info, and request human review; experts refine source allow-list and triggers. *Escalation:* If repeated loops occur (guidance keeps requesting same confidential info and control keeps intervening), notify IT/security and process owner after a set iteration threshold.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Demand-forecast support Agent	Pre-filters incoming customer requests; generates multiple demand scenarios (instead of one forecast) and checks feasibility under supply constraints; aligns with Sales; recommends planning scenarios to Supply Chain; escalates on disruption signals or recurring issues.	Ask about customer needs; access customer/order data; generate/compare scenarios; flag demand–supply mismatches; pass low-demand signals to Marketing Agent for potential promotions.	Modify confirmed orders/delivery components; commit to service levels without human approval.	*Rollback:* Hard-coded control gate requires explicit human approval (Sales + Supply Chain) before scenario recommendations can change operative plans; repeated bias triggers forecasting-team review of assumptions/models. *Escalation:* Notify Sales and Supply Chain on critical feasibility gaps; escalate to IT/AI governance if filters trigger unusually often (misuse/drift) or structural forecasting issues persist.
Insights Agent	Aggregates multi-dashboard and multi-source information; highlights relevant KPIs and bottlenecks; provides structured action recommendations/playbook steps; designed for downstream agent or human execution; escalates on missing/inconsistent data.	Query SQL/data warehouse; semantic and KPI-based search across dashboards/docs; generate structured insights and action items for humans/-downstream agents.	Execute high-impact actions autonomously (e.g., large schedule changes, financial decisions); access out-of-scope confidential data.	*Rollback:* Control agent monitors sources/-queries and halts on out-of-scope access or contradictions; BI/process owners validate and refine queries, permissions, and templates. *Escalation:* If repeated loops of conflicting recommendations or out-of-scope retrieval occur, control agent notifies BI and managers after an iteration threshold for remediation.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Forecast-monitoring Agent	Compares new forecasts vs. prior periods; cross-checks with inventory and shared supplier signals; computes deviations and combined risk indicators jointly with Demand-forecast support and Insights Agents; informs Demand Planning.	Access internal planning data, forecast histories, and inventory; ingest supplier-availability signals when shared; compute deviations and combined risk indicators with other agents.	Access supplier systems without agreements; unilaterally change production/procurement plans; override human decisions.	*Rollback:* Demand planners and SC managers manually adjust plans when thresholds/models cause wrong decisions; persistent errors trigger joint review by planning, analytics, and AI governance to recalibrate models/thresholds/integration. *Escalation:* Alert planners at early-warning limits or critical combined-risk levels; escalate to SC management and AI governance on repeated misalignments or major upstream constraints.
Waste-monitoring Agent	Samples BBD data; combines expiry, shelf-life history, prior waste, and upcoming demand curves; estimates waste risk and flags products/locations; escalates to Supply Chain and Order/Routing Agent when systemic risk emerges.	Collect/store BBD and shelf-life data; combine with forecasts and historical order patterns; estimate waste risk; flag items/locations requiring action.	Fully manage the system without human supervision; dispose of products or trigger large reallocations without approval.	*Rollback:* Secondary AI or rule-checker detects outliers/inconsistencies and can restart algorithm for suspected items; hub managers/inventory controllers correct data at source and adjust sampling logic. *Escalation:* Notify hub managers at local thresholds; escalate to Supply Chain and Order/Routing-management Agent on systemic overstocking/waste patterns.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Claims Agent	Retrieves and reconciles claim/product/transaction data across unconnected systems (SKU, order, batch/lot, invoice/-payment); flags unpaid invoices and discrepancies; estimates payment-delay likelihood; escalates on high volume/value or cash-flow gaps.	Retrieve/store claim data; match invoices, orders, and payments; flag unpaid invoices/discrepancies; estimate payment-delay likelihood/timing from historical patterns.	Modify financial records; commit/reverse payments; share confidential financial info externally without authorisation.	*Rollback:* Finance+IT review suspicious matches (SKU swaps, date formats), correct source data, update matching rules; systematic delay-misestimation triggers retraining/simplification under AI governance. *Escalation:* Notify AR/treasury for significant unpaid volumes or high-value discrepancies; escalate to finance management/risk committees for systemic issues; escalate to sustainability/circularity teams when data can support end-of-life decisions.
Maintenance-planning Agent	Extends capacity-risk monitoring with condition/failure data to propose preventive maintenance windows that minimise disruption while protecting machine health.	Analyse utilisation and condition-monitoring data; propose preventive windows; estimate maintenance risk and impact on capacity.	Override mandatory safety maintenance; shut down critical equipment without explicit human approval.	*Rollback/Escalation:* Maintenance planners validate/adjust windows; repeated mismatches trigger joint review with production, maintenance, and AI governance.
Bottleneck and action-taking Agent	Consumes signals from capacity-risk, forecast-monitoring, and insights agents to detect persistent bottlenecks and propose mitigation actions; enables semi-automation only for low-impact actions under strict rules.	Rank bottlenecks; simulate alternative routings/allocations; generate recommended actions and implementation plans.	Implement high-impact changes (e.g., major schedule revisions) without human approval; alter contractual service levels.	*Rollback/Escalation:* Ops managers validate feasibility and monitor realised effects; repeated underperformance triggers recalibration of simulation assumptions and thresholds.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Quality-management documentation / defect-tracing Agent	Automates quality documentation across inspection points and traces defect patterns over time (batches, suppliers, process steps) to support root-cause analysis and traceability.	Collect quality data; link defects to batches/suppliers/steps; produce traceability reports.	Modify quality records; downgrade critical defect flags.	*Rollback/Escalation:* Quality managers validate thresholds and sampling; systematic misclassification/missed defects triggers joint review with data engineering and quality teams.
Supplier-capability mapping Agent	Maps supplier capabilities (technology, capacity, lead time, sustainability attributes) to support supply chain reconfiguration during disruptions or technology shifts.	Aggregate supplier info from internal/external sources; classify capabilities; suggest alternative supplier options.	Change approved-supplier lists; sign/terminate supplier contracts.	*Rollback/Escalation:* Procurement validates suggestions and updates master data where appropriate; non-compliant recommendations feed model refinement.
Reimbursement and returns Agent	Operationalises returns and reimbursements using structured cases from the Feedback Agent, under strict controls and approvals.	Create return orders; prepare reimbursement proposals; track case status; check eligibility against policies.	Authorise payments beyond thresholds; bypass mandatory human approvals for high-value/exceptional cases.	*Rollback/Escalation:* Customer support and finance verify sampled/high-risk cases; recurrent eligibility errors trigger rule/policy updates.
Resourceand skill-allocation Agent	Recommends task-to-people assignments based on competence and availability, leveraging expert knowledge and HR/skills databases; highlights skill gaps and training opportunities.	Recommend staffing options; identify skill gaps; propose training opportunities.	Unilaterally reassign staff; modify contracts or working conditions.	*Rollback/Escalation:* Line managers approve/reject and provide feedback to refine skill models over time.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Training and coaching Agent	Provides targeted training/coaching content and contextual tips for complex tasks, informed by patterns captured by expert guidance; logs usage for learning analytics.	Generate training paths; push contextual tips; log usage for learning analytics.	Evaluate employees in formal HR processes without human oversight.	*Rollback/Escalation:* L&D teams monitor effectiveness and adjust content/targeting.
Marketing and promotion Agent	Triggered by demand-scenario signals, proposes promotions/discounts and simulates inventory/margin impacts within marketing governance constraints.	Propose promotions/discounts; estimate uplift; simulate impact on inventory and margins.	Launch campaigns without marketing approval; set prices outside defined corridors.	*Rollback/Escalation:* Marketing reviews campaign performance and tightens/relaxes rules accordingly.
Orderand routing-management Agent for waste minimisation	Uses waste-risk signals to propose prioritisation, re-routing, and stock transfers to reduce expiry-driven waste, subject to operational and contractual constraints.	Propose re-routing near-expiry items; recommend priority shipments; suggest hub-to-hub stock transfers.	Execute large-scale re-routing without human validation; violate regulatory/contractual constraints.	*Rollback/Escalation:* Supply Chain managers assess trade-offs vs. cost/service; misaligned recommendations lead to updated optimisation rules.
Circularity / end-of-life decision-support Agent	Uses rich cross-process data (claims, quality, usage, feedback) to propose compliant end-of-life options: reuse, refurbish, recycle, dispose.	Classify products by condition/reuse potential; propose end-of-life options compliant with policy and regulation.	Dispose of hazardous materials without specialist approval; ignore regulatory requirements.	*Rollback/Escalation:* Sustainability and compliance validate high-impact recommendations and update decision rules as frameworks change.
Predictive claims and payment-risk Agent	Extends claims handling with predictive risk scoring for claims and payment delays to anticipate cash-flow risk and prioritise follow-ups.	Analyse historical claim/payment data; score transactions by risk; propose prioritisation of follow-up actions.	Modify payment terms; autonomously contact customers/suppliers on sensitive financial issues without agreed protocols.	*Rollback/Escalation:* Finance and risk teams validate scores; systematic bias/inaccuracy triggers model adjustment or simplification.

Continued on next page.

Table C.13 continued.

Agent	Core function	Always allowed	Never allowed	Rollback path & escalation
Regulation and external-constraint monitoring Agent	Monitors trusted regulatory/news sources to map external constraints (trade restrictions, shutdowns, changing rules) onto operations and flag impacted routes/suppliers/customers.	Monitor trusted sources; match rules to operations; flag affected routes, suppliers, and customers.	Provide legal advice; commit the firm to interpretations without legal review.	*Rollback/Escalation:* Legal/compliance reviews and approves interpretations/actions; misinterpretations trigger refined rule mappings and source whitelists.

C.6. Supporting Literature Cross-Domain Analogy

The cross-industry analogical reasoning methodology demonstrates that effective translation depends on mapping relational structures rather than surface features. Interviews with AI and supply chain experts indicate that these structures should support graduated autonomy according to risk, distinguish monitoring from action, incorporate explicit fallback modes, ensure transparency for multiple stakeholders, and establish feedback loops that convert operational corrections into systematic improvements.

Airbus has addressed comparable governance challenges through extensive refinement, incident investigation, and certification over several decades. This analysis incorporates a technical description of the automation architecture and interviews with an Airbus line pilot and an AI/automation engineer, illustrating how these mechanisms function under operational pressure and how informal practices emerge when formal structures confront real-world complexity. The Airbus Automatic Flight System provides empirical evidence of the mature governance relations identified by the methodology.

Within Airbus terminology, the Automatic Flight System (AFS) encompasses the entire automation architecture responsible for managing the aircraft's flight path and engine thrust. Operationally, automation comprises the Flight Management and Guidance System (FMGS), the autopilot/flight director (AP/FD), the autothrust (A/THR) system, and the fly-by-wire (FBW) control laws (Federal Aviation Administration, 2022; Lelaie, 2012). Collectively, these systems decrease pilot workload by automatically controlling flight path and thrust, while continuously providing flight-envelope protection (Airbus, 2025b).

The Airbus Technical Training Manual – ATA 22 Auto Flight (Airbus, 2008) identifies four principal functional domains of the Automatic Flight System (AFS), as illustrated in Fig. C.1:

- Flight Envelope: continuously monitors airspeed, aerodynamic forces, and the angle of attack, maintaining the aircraft within certified safety limits by reducing or blocking control inputs that would exceed these boundaries.
- Flight Guidance: comprises the autopilot (AP), flight director (FD), and autothrust (A/THR) systems, which direct the aircraft along the intended flight path.
- Flight Management: managed by the Flight Management and Guidance System (FMGS), which is responsible for flight planning, navigation, and performance calculations.
- Maintenance and Test Facilities: incorporate built-in test equipment and ground or flight test routines that record faults, generate post-flight reports, and support fault isolation, thereby ensuring AFS integrity, configuration control, and reliability.

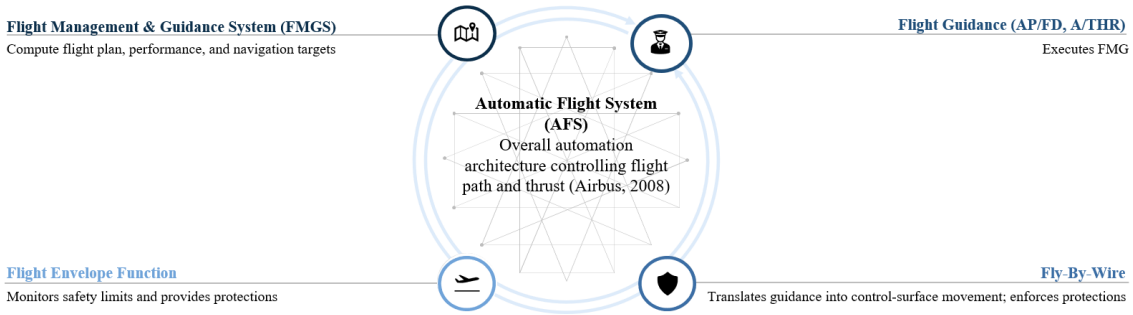


Figure C.1: Hierarchy of Airbus automation subsystems within the AFS.

Flight Management and Guidance System

The Flight Management and Guidance System (FMGS) serves as the computational core of the Automatic Flight System (AFS). It calculates both lateral and vertical flight paths, determines target speeds and altitudes, and manages the navigation database. Two redundant Flight Management and Guidance Computers (FMGCs) operate in parallel and cross-monitor each other to enhance reliability and facilitate fault detection (Airbus, 2018; Federal Aviation Administration, 2022; Lelaie, 2012). The FMGS interfaces with pilots via the Flight Control Unit (FCU) and Multipurpose Control and Display Units (MCDUs), and operates in two primary modes:

- Managed mode: the system automatically follows targets (speed, altitude, path) computed from the flight plan;
- Selected mode: the pilot inputs parameters directly via the FCU or MCDU.

This dual-mode architecture maintains a balance between automation and pilot authority, thereby supporting situational awareness (Airbus, 2025b; SKYbrary, 2025).

Flight Guidance

The Flight Guidance function converts outputs from the Flight Management and Guidance System (FMGS) into control actions via the autopilot (AP), flight director (FD), and autothrust (A/THR). These systems collectively maintain aircraft trajectory, altitude, and speed, thereby reducing pilot workload (First, 2021). Airbus operational guidance, supported by accident and incident investigations such as ATSB report AO-2016-012, recommends engaging the autopilot during instrument approaches to minimize workload and ensure managed guidance (Australian Transport Safety Bureau, 2018). A notable aspect of Airbus design is the use of ”non-moving” thrust levers, which remain in fixed detents (such as takeoff/go-around, climb, and idle), while autothrust electronically regulates engine output. Pilots are specifically trained to monitor autothrust engagement, lock warnings, and mode transitions (Airbus, 2008).

Two features at the human–automation interface serve as critical safeguards. The Flight Mode Annunciator clearly displays active and armed modes, as well as any changes in flight control laws, to the crew. Additionally, the Flight Control Unit (FCU) push/pull logic enables a rapid transition from managed guidance to direct selection of specific targets, such as heading, vertical speed, altitude, and airspeed, when predictability or data integrity is uncertain (Airbus, 1998; Goteman & Dekker, 2006).

Fly-By-Wire System

Beneath the guidance layer, the Fly-By-Wire (FBW) system executes automation commands generated by the Automatic Flight System (AFS). Instead of traditional mechanical linkages, the FBW system employs electronic control to transmit both pilot and autopilot inputs, enforcing flight-envelope protection and fault isolation while

integrating pilot authority with automation. This integration is central to Airbus' philosophy of maintaining human supervision over automated flight control (Brière et al., 1995; Goupil, 2011b; Traverse et al., 2006).

Control Laws and Graceful Degradation

Airbus specifies three control laws that degrade progressively in response to system failures, thereby maintaining aircraft controllability.

- Normal Law: delivers comprehensive flight-envelope protection, including constraints on pitch, load factor, and bank angle. These protections serve as boundaries for aircraft response while faults are detected and isolated (Malinge, 2016).
- Alternate Law: activated following specific failures; some protections are lost, but stability augmentation is retained. Upon transition to Alternate Law, the autopilot and flight director disengage, which increases crew workload and necessitates enhanced monitoring (P. S. D. Airbus, 2019).
- Direct Law: a fallback mode in which pilot inputs are relayed directly to the control surfaces without augmentation.

Pilot training emphasizes understanding system transitions between modes and recognizing associated messages on the Electronic Centralized Aircraft Monitor, as well as the corresponding handling requirements (SKYbrary, 2025). Incidents such as Air France Flight 447 highlight the significance of this progression and the necessity of maintaining a consistent pitch-and-thrust response when protections are reduced (d'Enquêtes et d'Analyses, 2012).

Redundancy and Safety Architecture

Airbus integrates redundancy and dissimilarity within both hardware and software architectures. Distinct processor types and code bases are deployed across computer classes, and three independent sensor channels are used for air data and inertial inputs. Furthermore, electrical and hydraulic power systems are separated to prevent any single fault from compromising the entire system. These strategies are supported by formal fault detection and isolation procedures, followed by graceful reconfiguration, which together ensure continued aircraft controllability as the system adapts (Brière et al., 1995; Goupil, 2011a). Continuous data integrity verification and system monitoring are essential components of Airbus operations. Three independent Air Data and Inertial Reference Units are cross-validated using majority voting, and built-in test equipment compares control and monitor channels with actuator feedback to identify faults. These practices are fundamental to Airbus's dependability program (Brière et al., 2001; Goupil, 2011a). These patterns of diverse redundancy, voting and cross-checks, graceful degradation, explicit mode displays, and a straightforward fallback interface are directly applicable to resilient AI systems. They constrain failures, maintain a safe minimum level of service, and ensure human oversight when data are unreliable or components malfunction. Collectively, these functions constitute a layered control architecture. The Flight Management and Guidance System computes targets, Flight Guidance translates these targets into commands, Fly-By-Wire executes the commands while enforcing protections, and the Flight Envelope function supervises operational limits.

How Airbus Tests Automation in Rare Scenarios

Ensuring that automation systems can address unexpected events, including those with very low probability, is essential. In response, Airbus has adopted a data-centric paradigm to manage the increasing complexity of flight operations (Kilic et al., 2024; Wijnands et al., 2024). Due to the logistical, financial, and safety challenges associated with collecting sufficient real-world flight data, the industry now relies on synthetic data generation, automated scenario modeling, and digital twin technology (Chigot et al., 2025; Sprockhoff et al., 2024; Stefani et al., 2025). This methodology allows manufacturers to "build each aircraft twice," first within a digital environment and subsequently in physical form, ensuring that automated systems are thoroughly matured through comprehensive, data-driven simulation throughout the product lifecycle (Airbus, 2025a).

A design philosophy emphasizing redundancy and dissimilarity is employed to maintain automation continuity, particularly during primary sensor failures. According to an Airbus AI and automation expert, ”extreme or rare situations in safety-critical flight control are addressed using the same assurance logic as for non-AI systems: redundancy and dissimilarity at the design stage, followed by validation through extensive testing or formal proof.” This strategy entails developing systems that utilize fundamentally different methods to estimate identical state variables. For example, synthetic air data concepts based on analytical redundancy combine dissimilar sources and models to recover estimates when direct sensing is unavailable (Kilic et al., 2024; Miltner et al., 2014). The expert further noted that ”robustness testing, including targeted synthetic failure scenarios, is a standard practice for validating equipment-failure behavior,” with extremely rare events primarily validated through synthetic data that systematically addresses conditions not adequately covered by real-flight data.

Synthetic environments enable comprehensive coverage across multiple validation domains. Flight simulators produce extensive image datasets with systematically controlled variables for training deep vision models used in autonomous perception tasks, such as runway detection (Chigot et al., 2025). These simulators capture undersampled edge cases like rare illumination and visibility conditions that are difficult or costly to obtain in real operations. Similarly, synthetic trajectory generation broadens the coverage of landing and approach behaviors (Wijnands et al., 2024). Testing of AI-based and highly automated functions is organized around a clearly defined Operational Design Domain (ODD), which specifies the conditions under which the system is intended to operate and outside which performance guarantees are not provided (Stefani et al., 2025). Automated scenario generation facilitates systematic exploration of both logical and concrete scenarios at scale, enabling the creation of variations that are impractical to reproduce in real flight tests (Sprockhoff et al., 2024; Stefani et al., 2025). This comprehensive coverage approach uncovers discontinuities, unstable advisories, or boundary-condition failures that are not present in nominal datasets (Stefani et al., 2025). Regarding sufficiency of coverage, the Airbus expert indicated that catastrophic cases not covered by testing should have a probability of occurrence no greater than:

$$P_{\text{failure}} \leq 10^{-9} \text{ per flight hour.} \tag{C.1}$$

Satisfying this requirement necessitates the integration of large-scale synthetic scenario testing with robust verification arguments, including test evidence and formal methods that align with certification standards.

The development lifecycle progresses systematically from fully virtual environments to hybrid physical–digital integration. A digital twin, defined as a virtual model that maintains meaningful correspondence with a physical asset, supports analysis, testing, and ongoing improvement throughout this process (Boyes & Watson, 2022). Airbus utilizes digital twins to accelerate development by advancing validation into high-fidelity virtual environments, conducting virtual flight-test campaigns, and connecting simulators with avionics and system integration environments for early end-to-end validation (Airbus, 2022, 2025a). This methodology is increasingly recognized as a pathway to ”virtual certification,” where evidence is generated through extensive simulation prior to flight, resulting in efficiency gains such as reduced reliance on costly flight-test hours (Aydemir et al., 2025). The validation process culminates in physical–digital integration rigs, such as the ”Iron Bird,” which incorporate representative hydraulics, electrics, and flight controls to evaluate integrated behavior under controlled conditions (Airbus, 2017). Recent research situates these rigs along a continuum ”from Iron Birds to Digital Twins,” with hybrid infrastructures ensuring continuity between virtual evidence and physical integration (Aydemir et al., 2025). Flight simulators further extend this ecosystem into training and operational readiness, offering capabilities advanced enough to reduce reliance on aircraft availability for qualification and recurrent training (Airbus, 2025c). Across these layers—synthetic data, scenario generation, digital twins, hybrid rigs, and full-flight simulators—the primary objective is to expand evidence and assurance for automation behavior in rare, safety-critical conditions, while managing cost and risk through systematic, high-fidelity simulation (Airbus, 2025a; Aydemir et al., 2025).

C.7. Interview - Airbus Pilot

Interview Summary

Automation is the standard way to fly an Airbus. ECAM (Electronic Centralised Aircraft Monitoring) detects abnormal situations, highlights the main failure first, and suppresses less important alerts using colour coding and phase-based inhibition. In routine operations, the autopilot is engaged shortly after take-off. When entering data into the flight computer would take too long–especially in busy airspace–the crew use the control knobs or brief hand-flying to act quickly, then return to full automation once the aircraft is stable and within limits. As shown by the Air France Atlantic case, if probe data are corrupted and the autopilot disconnects, the automation cannot reliably cross-check and disengages; the crew then take manual control and verify which indications are trustworthy.

In failure situations, the crew first stabilise the aircraft with simple pitch and power, and follow memorised procedures without improvising. Redundant systems and regular simulator training help keep recovery times short. If automated guidance seems doubtful, pilots cross-check raw data–thrust, temperatures, attitude and colour arcs–and consult the manuals. Before the approach, they brief possible degradations and consider diverting if needed. Known limits for wind, weight and runway length shape the plan, often with support from Flight Dispatch. The guiding principle is pace and clarity: if flight-computer entry is too slow or guidance is unclear, use the knobs or brief hand-flying to respond quickly, then restore full automation once stabilised.

Open Coding

Table C.14: Open Coding Analysis - Airbus Pilot

Code Name	Definition	Source Extract
ECAM cue	ECAM gives the first and authoritative abnormal cue	“The aircraft’s behaviour is always codified by the ECAM… which provides warnings…”
Priority order	ECAM orders warnings so the main failure is addressed first	“…provides warnings with priorities and logic… so that you address the main failure…”
Central monitor	One central system monitors all connected systems and reports failures	“…this centralised system, which monitors all connected systems and reports failures…”
Simultaneous faults	Handling of multiple failures at once	“…reports failures–especially if they are simultaneous…”
Label message	Failures appear as a specific labelled message	“…informed by the ‘FAILURE (VARIA)’ message…”
Real failure	“Real” failure is the one that threatens normal control	“The real failure is the one that jeopardises normal control of the aircraft.”
ECAM first	System warning arrives before aural or feel cues	“…the ECAM–arrives immediately, and you see it before you even hear it.”
Yaw cue	Engine failure recognised by yaw from thrust asymmetry	“…engine failure… you see a yaw due to thrust asymmetry…”
Auto by design	Automation is the intended, normal management mode	“You generally use the automation because the aircraft is designed to be managed with automation…”

Continues on next page.

Table C.14 continued.

Code Name	Definition	Source Extract
Computer guidance	The system proposes steps and corrective actions	“Usually, the computer itself gives you the steps to follow and the corrective actions.”
Semi auto mode	Use of modes that are not fully manual to simplify control	“You can choose functions that are not fully manual–semi-automatic…”
Experience use	Greater experience enables better semi-auto use	“You do this mostly when you have a certain experience…”
Selected functions	Pilot-selected settings with automation yield better behaviour	“…by setting functions you select… they deliver better behaviour.”
Early AP	Autopilot engaged soon after lift-off in normal ops	“…after five seconds you can engage the autopilot.”
Rotation rate	Target rotation rate to pitch target	“You rotate at 3° per second to 15°.”
Takeoff thrust set	Take-off thrust is pre-set before AP engagement	“You have already set the necessary thrust for take-off.”
Simple actions first	In failures, perform simple stabilising actions before AP	“…you first have to perform simple actions that you study…”
AP drop causes	AP may drop out due to missing data or manual actions	“…may have disconnected because data were missing, or because you had to perform manual actions…”
Keep trajectory	Systems aim never to abandon the flight path	“…it never abandons the flight trajectory…”
Redundant channels	Multiple channels provide redundancy for stability	“…built this way and is redundant (because there is more than one channel)…”
Action priorities	System prioritises actions to take	“The computer itself gives you the priorities of the actions to be taken…”
Reengage or manual	After prioritisation, re-engage AP or fly manually	“…either re-engage automation or manage the aircraft manually…”
Correct sequence	Emphasis on correct order of actions	“…with the correct actions in the correct order.”
Full auto busy	In crowded airspace, use maximum automation	“…in very crowded airspaces… you use all available automation.”
Workload shield	Automation reduces workload when attention is divided	“…lack of automation would increase your workload.”
Almost mandatory	High-density environments make full automation near-obligatory	“…it is almost mandatory… to use full automation…”
Knobs fast	Use knobs when typing would be slower	“…act on knobs that are more immediate.”
Temporary split	One pilot knobs, the other enters data	“…temporarily, while the other pilot enters what you want…”
Back to max auto	End-state preference is maximum automation	“…you always try to manage everything with maximum automation.”
AP drop conditions	AP can drop out with speed or turbulence variations	“It is possible for the autopilot to disconnect in… speed variation, turbulence…”

Continues on next page.

Table C.14 continued.

Code Name	Definition	Source Extract
Reengage when normal	Re-engage AP once parameters are within limits	"…if… parameters are back within limits, you can re-engage it…"
Memory items	Use memorised, standard sequences for cases	"…actions to perform from memory–memory items…"
No improvisation	Emergencies follow established procedures	"You don't improvise; these are emergency corrective actions…"
OEM and operator	Procedures are set by manufacturer and airline	"…established by both the manufacturer and the operator…"
Instinctive execution	Mastery leads to near-automatic performance	"…perform in an almost automatic, instinctive manner."
Entry error	Dubious guidance may stem from pilot data entry	"…could be due to an entry that you made."
Dual check	One pilot enters, the other verifies; briefed again	"…entered by one pilot, checked by the other… reiterated during the briefing."
Expected behaviour	Experience provides expected speed/thrust/attitude patterns	"…you know that the aircraft has certain behaviours…"
Raw checks	Use thrust, EGT, attitudes as basic cross-checks	"You look at… engine thrust… and exhaust temperature."
Amber prewarn	Colour change to amber can precede ECAM message	"…start to become amber… sometimes even before the… ECAM… warns you."
Use manuals	Use systems manuals if ECAM does not flag	"…in case of something dubious not identified by the ECAM, you use the manuals."
Find chapter	Retrieve the relevant systems chapter for checks	"…by retrieving the relevant systems chapter…"
Standard brief	Normal approach brief covers procedures and context	"…you give a normal briefing… covering procedures…"
AP disconnect brief	Specify when the autopilot will disconnect	"…specify… when the autopilot will disconnect…"
Autobrake choice	May choose to avoid autobrake to manage exit	"…decide not to use the autobrake… to manage the exit better."
Emphasised brief	Briefing is strengthened when a failure exists	"…a briefing is emphasised… when there is already a failure."
Anticipate further fail	Discuss what happens if another failure occurs	"…anticipate what could happen in case of a further failure…"
Limits and diversion	Consider limits and possible diversion	"…limitations for landing could even entail a diversion…"
Known limits change	Pre-existing failures change destination or alternates	"…known limitations… can change the choice of destination airport…"
Weather margins	Small weather changes can force diversion	"…with a small worsening of the weather… divert…"
Fuel for diversion	Adjust fuel strategy to ease diversion	"…a fuel issue is managed differently to make a diversion easier."

Continues on next page.

Table C.14 continued.

Code Name	Definition	Source Extract
Limits wind weight runway	Failures impose wind, weight, runway limits	"…could lead to wind… landing-weight… or runway limitations…"
Runway contamination	Contamination can preclude landing	"(e.g., if it is contaminated, precluding landing)."
Dispatch support	Dispatch helps and adds extra margins	"…Flight Dispatch… who take additional margins."
Typing latency trigger	Simplification triggered by slow keypad entry	"The trigger is recognising that manual digital entry is too slow."
Knobs immediacy	Knobs provide immediate value selection	"…with knobs… you immediately set values."
Keep task order	Avoid losing the sequence of other tasks	"…not to lose the sequence of other actions…"
Less typing low alt	Reduce FMS typing at low altitude	"…at low altitudes, you tend to use digital entries less."
Disconnect AP	Option to hand-fly when execution speed lags	"…possibility of disconnecting the autopilot and… manual piloting."
Daily pattern	This pattern occurs daily in approach/landing	"…approaches and landings that happen daily."
Alert severity	Alerts scaled by severity to avoid undue disturbance	"…designed and codified according to their severity…"
Silence option	Some alerts can be silenced	"…possibility of silencing them."
Emergency tone	Emergencies use strong, repetitive tones	"We move from repetitive and strong audible warnings…"
Info no action	Informative alerts require no action	"…informative warnings of a failure that does not require anything of you."
Short info tone	Informative alerts use simple, short sounds	"This informative warning uses a simple, short sound…"
Phase inhibition	Alerts can be fully inhibited in critical phases	"…warnings are completely inhibited… in some flight phases…"
Post phase reactivate	Alerts re-activate after the critical phase	"…when that phase ends, they are immediately reactivated…"
Status bar	A persistent status or memo area exists	"A status bar exists. It shows… memo messages…"
Ops and faults shown	Bar shows operational status and failures	"…operational status indications and failure indications."
Green normal	Green indicates normal condition by phase	"Messages in green indicate normality… by phase."
Hazard levels	Display changes with hazard or informative level	"…different condition depending on the level of hazard…"
Brake colour example	Example of colour semantics by phase	"…parking brake… amber during engine start… red in flight…"

Continues on next page.

Table C.14 continued.

Code Name	Definition	Source Extract
Reminder messages	Reminder items like seat belts shown	"…something you need to remember, like passenger seat belts…"
Stability edge cases	Stability may be slightly lost in specific cases	"…slightly lost its stability… during take-off… engine failure."
Sim training	Regular simulator repetition of such manoeuvres	"…repeating them multiple times in the simulator…"
Quick recovery	Training yields quick, instinctive reactions	"…reactions are quick and instinctive."
Limited recovery time	Overall time to recover is limited	"…time required for recovery is quite limited."
Weather time driver	Weather-related disturbances can consume time	"…related to weather disturbances… simulated mostly…"
Multiple failures	Multiple failures can affect indications or automation	"…multiple simultaneous failures."
Procedure updates	Procedures updated during aircraft development	"…procedures are updated and refined."
Joint governance	Airbus and airlines jointly analyse and refine	"…by the manufacturer… with the help of all the airlines…"
Not single pilot	Changes are not decided by individual pilots	"The change is not decided by a single pilot."
Tech analysis	Technical departments analyse parameter excursions	"A technical department analyses what happened…"
OEM feedback loop	Airlines suggest changes; Airbus implements updates	"…suggested to Airbus… implemented by Airbus…"
AP disconnect cue	Autopilot disconnects when reference data become invalid	"Turbulence corrupted probe data… the autopilot disconnected."
Discrimination loss	Automation cannot reliably choose between conflicting inputs	"…automation could no longer discriminate reliably…"
Manual reversion	Crew reverts to manual control and validates trustworthy indications	"…the crew reverted to manual control while trying to identify trustworthy indications."
Multi-failure limit	Multiple concurrent failures can exceed computer assistance capability	"…when there is more than one failure, the computer can no longer assist."
Pace–clarity trigger	Take control when tasks are time-critical or automation is ambiguous	"The decision to take control is driven by pace and clarity… if automation is degraded or ambiguous, I hand-fly…"
Knobs over keypad	Use knob-based commands when FMS entry is too slow	"…if… entry is too slow, I switch to knob-based commands for a faster response."

Continues on next page.

Table C.14 continued.

Code Name	Definition	Source Extract
Instrument-guided hand-fly	Hand-fly using instrument guidance, with the final segment fully by hand	"…other instrument functions still provide guidance… with only the final segment being visual."
Procedure sequence	ECAM triggers ordered steps that must be followed exactly	"The warnings trigger sequences that must be performed exactly in order."

Themes and Supporting Open Coding

Table C.15: Themes and Supporting Open Coding Names – Airbus Pilot Interview

Theme Name	Supporting Open Coding Name
Theme 1: ECAM-centred monitoring and graded alerting	ECAM cue; Priority order; Central monitor; Simultaneous faults; Real failure; Label message; ECAM first; Alert severity; Emergency tone; Info no action; Short info tone; Phase inhibition; Post phase reactivate; Status bar; Ops and faults shown; Green normal; Hazard levels; Brake colour example; Reminder messages
Theme 2: Automation by default and layered control authority	Auto by design; Computer guidance; Early AP; Takeoff thrust set; Rotation rate; Keep trajectory; Redundant channels; Action priorities; Reengage or manual; Full auto busy; Workload shield; Almost mandatory; AP drop causes; AP drop conditions; Reengage when normal; Stability edge cases; AP disconnect cue; Discrimination loss; Manual reversion; Multi-failure limit; Instrument-guided hand-fly
Theme 3: Tactical simplification via knobs, semi-auto modes and reduced typing	Semi auto mode; Experience use; Selected functions; Knobs fast; Temporary split; Typing latency trigger; Knobs immediacy; Keep task order; Less typing low alt; Disconnect AP; Knobs over keypad; Daily pattern; Pace–clarity trigger; Instrument-guided hand-fly
Theme 4: Proceduralisation, briefing and governance of responses	Memory items; No improvisation; OEM and operator; Instinctive execution; Standard brief; AP disconnect brief; Autobrake choice; Emphasised brief; Anticipate further fail; Limits and diversion; Known limits change; Weather margins; Fuel for diversion; Limits wind weight runway; Runway contamination; Dispatch support; Sim training; Weather time driver; Multiple failures; Procedure updates; Joint governance; Not single pilot; Tech analysis; OEM feedback loop; Procedure sequence
Theme 5: Cross-checking, plausibility checks and manual verification of automation	Entry error; Dual check; Expected behaviour; Raw checks; Amber prewarn; Use manuals; Find chapter; Yaw cue; Real failure; ECAM first; AP disconnect cue; Manual reversion

C.8. Interview - AI and Automation Engineer

Interview Summary

The interview draws a clear boundary between classic, certifiable automation and modern AI/ML, and explains why Airbus largely keeps ML out of safety-critical control today. Airbus' position is that flight-critical systems (DAL-A) remain a "traditional rules + certifiable code" domain because AI certification is not yet standardised at the level

aviation requires. The only “AI” accepted inside critical functions is described as very simple and fully testable, essentially functioning like a compressed lookup table: an AI model can represent a very large mapping from inputs to outputs, but Airbus still requires exhaustive coverage of the input space to prove correct behaviour. An example mentioned is the A350 braking system, where this type of fully testable AI can be used without introducing opaque decision-making. Where AI is used today depends on criticality. For critical functions, AI is mostly used upstream as a development accelerator (e.g., helping generate code that is then certified through formal methods or exhaustive testing), while the onboard software remains deterministic. For non-critical functions, Airbus can embed AI models directly, with the interview citing computer vision for runway-incursion/obstacle detection as a representative case. Even there, the AI is positioned as advisory—it can raise an alert, but it does not take control or autonomously execute manoeuvres; responsibility stays with the pilot. A central theme is how Airbus manages faults: not via “AI self-healing,” but through architecture, redundancy, and controlled degradation. The interview highlights a command/monitor architecture: one computer “commands,” a separate monitoring computer independently checks and has authority to shut down the command unit if outputs disagree. When faults are detected, the system isolates the faulty unit, reconfigures to a redundant channel, and often transitions into a degraded mode with simpler control laws. This prevents a single error (sensor or computer) from propagating across safety-critical functions—an explicit parallel to avoiding cascade failures in supply chains. The interview contrasts this with a common Boeing pattern of triple modular redundancy with voting, noting both approaches aim to ensure one failure cannot become systemic. On the concern of AI “hallucination,” Airbus’ “hard safety layer” is framed as certification constraints plus design choice: if Airbus cannot prove—via exhaustive testing or formal proof—that a component meets extremely low failure-probability expectations (the interview references the 10□□ order of magnitude), then it is not allowed into the critical control loop. Consequently, Airbus does not permit complex, opaque ML models to directly command critical flight controls under today’s certification regime. Human authority is treated as non-negotiable. The interview states Airbus’ philosophy is that the pilot has the last word, with clear ways to disconnect automation, trigger degraded modes, or revert to full manual control. When faults occur, the pilot is informed primarily in terms of system failure/degraded mode and associated procedures, not deep technical diagnosis. Detailed root-cause analysis is done after the flight using logs, and the interview notes that issues often trace back to specification gaps rather than simple component failures—leading to updates in requirements and certification artefacts. The interview is also explicit that online learning is not allowed in safety-critical flight functions: behaviour is frozen, validated, and certified before operation. Any learning or retraining for non-critical AI (e.g., improving performance in new conditions like snow) happens offline, after engineers analyse errors, update training data, and re-validate before deployment. Finally, the interview stresses traceability (“digital black box”) as essential for significant faults and anomalies: systems generate logs to reconstruct what happened and why, mainly to support post-event engineering analysis rather than to provide pilots with detailed explanations in real time. On pilot skill erosion, the respondent recognises that automation dominates cruise, but argues that take-off and landing remain engagement-intensive, and training/procedures are used to ensure pilots retain the ability to take over when automation disconnects or degrades

Open Coding

Table C.16: Open Coding Analysis - Automation and AI Engineer

Code Name	Definition	Source Extract
Certification limits	Certification for learning based safety-critical systems is not yet standardised, so Airbus remains in a traditional, rule-based, certifiable-software paradigm.	“For safety-critical systems, we are still essentially in a “traditional rules + certifiable code” world. The certification of AI-based systems is not yet standardised, and current methods are still in an early phase.”

Continues on next page.

Table C.16 continued.

Code Name	Definition	Source Extract
Compressed lookup tables	Use of simple, fully testable learning models as compact representations of large lookup tables, where all input–output combinations can be exhaustively verified.	"The only AI we can use today in critical functions is very simple and fully testable – essentially AI as a compressed lookup table. Instead of storing a huge table with millions of input combinations, we use an AI model to represent it, but we exhaustively test all input combinations to prove that the outputs are correct."
Critical control without learning	For critical flight control and similar systems, Airbus avoids complex learning models and relies on classic algorithms or very simple models that can be fully tested.	"So for flight control and other DAL-A critical systems, we are not deploying complex machine-learning models. We use either classic, pre-programmed algorithms, or very simple AI that can be completely tested over 100% of the input space."
Learning for code generation	In critical applications, learning systems are used off-board to generate code, while the final onboard implementation remains deterministic, certifiable software.	"For critical systems, AI's role today is upstream: we use AI as a tool to generate code that will then be certified (for example, using formal proofs or exhaustive testing of all branches), but the AI model itself is not embedded in the aircraft."
Advisory computer vision	Directly embedded learning models are confined to non-critical functions, where they provide advisory alerts (for example, computer vision for runway incursion detection) but do not take control of the aircraft.	"For non-critical functions, we can embed AI models directly. A typical example is computer vision for runway incursion detection. This AI would provide additional safety by raising an alert if it detects an obstacle on the runway. However, it does not take control of the aircraft; it is purely advisory/informative."
Assistive role of automation	Current learning systems mainly assist and advise, or support development, rather than autonomously flying and executing manoeuvres.	"So today, AI is mainly an assistive/advisory layer in non-critical functions, and a development tool for critical code, not an autonomous pilot that executes manoeuvres."
Redundant reconfiguration	Fault handling in critical systems relies on redundancy and automatic reconfiguration rather than self-healing logic based on learning.	"For critical systems, the primary mechanism is redundancy with automatic reconfiguration, not an AI "self-healing" logic."
Command and monitor pattern	Use of a command and monitor architecture where a monitoring computer checks the command unit and can shut it down if discrepancies are detected.	"Airbus uses a command/monitor architecture: One computer acts as the command unit. A separate monitoring computer performs the same function but only has the authority to shut down the command computer if it detects discrepancies or faults."

Continues on next page.

Table C.16 continued.

Code Name	Definition	Source Extract
Degraded modes with pilot	When faults occur, functions are transferred to redundant computers in degraded mode, after which the pilot decides how to continue and can take manual control.	"When a fault is detected, the faulty unit is disconnected and the function is automatically transferred to another redundant computer, often with a degraded mode using simpler control laws. The system first tries to reconfigure itself through redundancy and switch to a safe degraded mode. The pilot can then decide how to proceed and, if needed, take full manual control."
Diverse redundancy	Critical systems are protected by redundant command and monitor computers implemented with diverse hardware and software to minimise common-mode failures.	"At Airbus, for critical systems we use command + monitoring redundancy: A command computer controls the aircraft. A separate monitor computer, implemented in a different programming language and hardware, checks the output. By diversifying hardware and software between command and monitor, the probability that the same bug appears in both is very low. This prevents a single fault from propagating through the whole system."
Triple system voting	Boeing often applies triple-modular redundancy where three systems compute independently and a voter selects the majority output to reject faulty signals.	"Boeing, by contrast, often uses a triple-modular redundancy + voter approach: three systems compute independently, and a voting mechanism selects the majority output."
No cascading failures	Architectures are designed so that a single component failure is isolated and does not cascade into systemic failure across safety-critical systems.	"Both approaches are designed so that a single component failure does not cascade into a systemic failure."
No opaque models in control	Complex, opaque learning models are not allowed in critical roles unless failure probabilities below one in a billion can be formally demonstrated.	"For DAL-A critical functions, we do not allow complex, opaque AI models (e.g. large deep networks) in the control loop at all unless we can: fully cover the entire input space, or provide a formal mathematical proof or exhaustive testing guaranteeing a failure probability below 10^{-9} (one in a billion)."

Continues on next page.

Table C.16 continued.

Code Name	Definition	Source Extract
Limits on learning systems	Learning systems are restricted either to fully testable smart lookup tables in critical contexts or to offline development tools, never as an uncontrolled live decision-maker.	“In practice, this means: Critical control is handled by deterministic, certified code. AI is either: a fully testable “smart lookup table” (input space fully covered), or used offline as a development tool (e.g. code generation), not as a live decision-maker.”
Pilot final authority	Design philosophy ensures the pilot always has the final authority and can overrule automated systems.	“Yes. Airbus philosophy is that the pilot has the last word.”
Manual override options	Pilots can trigger degraded modes, switch to full manual control, and completely override automation whenever necessary.	“If something goes wrong: The system may automatically enter degraded mode and disconnect the faulty computer. The pilot can also trigger degraded modes or switch to full manual control. There is always a way to completely override the automated control.”
Design contrast with Boeing	Comparison with the Boeing situation is used to emphasise Airbus’ requirement that pilots can always take over and even override protections.	“This is a key design difference highlighted when comparing to the Boeing 737 MAX situation: in Airbus aircraft, the pilot can always take over control and even override certain protections (e.g. pitch limits) if necessary.”
Action focused feedback	In real time, pilots see clear failure indications and associated procedures, rather than detailed technical explanations of root causes.	“From the pilot’s point of view, they mainly see failures and associated procedures, not a deep technical root cause. … So: In real time, the focus is on clear indications and manageable actions, not full technical explanation.”
Post flight log analysis	Detailed fault analysis is performed by engineers on the ground using logs recorded during the event.	“When a fault occurs: It is logged by the system. The detailed analysis of why it failed is performed after the flight by engineers on the ground, using those logs.”
Specification gaps	Investigations often reveal that unexpected behaviour stems from specification gaps rather than pure component failures, leading to specification updates.	“Often, the true root cause turns out to be a specification gap (a case not foreseen) rather than a pure component failure, and the specifications are updated accordingly.”
Frozen critical behaviour	Safety-critical systems do not adapt or learn in real time; their behaviour is fixed and validated before operation.	“For safety-critical functions, systems are essentially frozen and validated before operation. They do not learn online during a flight.”

Continues on next page.

Table C.16 continued.

Code Name	Definition	Source Extract
Offline learning loop	Non-critical learning systems are updated offline by adding problematic cases to training data, retraining, and re-validating models before redeployment.	"For non-critical AI (e.g. a vision system that raises advisory alerts), learning happens offline: If the system generates a false alert or misses a case … The problematic case is added to the training data, and the model is retrained and revalidated before being deployed again."
Monitor shutdown and takeover	When conflicting outputs are detected in command and monitor setups, the monitor can shut down the command computer and let a redundant unit take over with simpler control laws.	"Airbus approach (command/monitor): A command computer and a monitor computer compute the same function independently. The monitor compares outputs; if it detects a discrepancy, it can shut down the command computer. Another redundant computer takes over, often with simpler degraded laws."
Voting conflict resolution	Triple-redundant architectures resolve conflicting data by majority voting among three independent systems.	"Boeing approach (voting): Three parallel systems compute independently. A voter compares them and selects the majority output; if one disagrees, it is effectively outvoted."
Fault isolation	Conflict handling is based on redundancy and isolation rules, not on a single model improvising reconciliations; pilots can always intervene.	"In both cases, the principle is redundancy + fault isolation, not "let a single model improvise a reconciliation". The system tries to resolve the conflict internally via redundancy and architectural rules, and the pilot can always take over if the situation requires."
Flight phase roles	Flight is divided into phases where some (take-off, approach, landing) require active pilot flying, while cruise allows extensive autopilot use.	"Operationally, the flight is divided into phases: Take-off and landing: phases where the pilot is actively flying and must be fully engaged. Cruise: roughly 90% of the flight, where the autopilot can handle almost everything and the pilot has relatively little to do."
Preserve manual skills	System and training design aim to keep pilots able to take over at any time and maintain manual flying skills despite high automation in cruise.	"The system is designed so that critical phases (take-off, approach, landing) still require active pilot control and attention, which naturally helps maintain skills. … the philosophy remains that the pilot must be able to take over at any time, and training is organised to preserve manual flying proficiency."

Continues on next page.

Table C.16 continued.

Code Name	Definition	Source Extract
Traceability for anomalies	Significant faults or anomalies are logged to enable reconstruction of events and diagnosis of causes.	"Traceability is very important, especially when something goes wrong. For any significant fault or anomaly: The system generates logs that record what happened. These logs are later used by engineers to reconstruct the event, understand the cause …"
Logs for design updates	Post-event analysis of logs informs updates to specifications and system design, closing the feedback loop between operation and certification.	"These logs are later used by engineers to … understand the cause (component failure, software bug, specification gap, unexpected condition), and update specifications or design if needed."
Selective traceability	Deep traceability is primarily required for critical faults and anomalies rather than every minor automated action in real time.	"Pilots themselves usually cannot see the detailed "why"; they see that something failed and follow operational procedures. The deep traceability and explanation are mainly for post-event analysis on the ground, not for every minor automated action in real time. For critical functions, this traceability and ability to analyse and learn from any issue is essential."

Themes and Supporting Open Coding

Table C.17: Themes and Supporting Open Coding Names – AI and Automation Engineer Airbus

Theme Name	Supporting Open Coding Name
Theme 1: Conservative use of learning systems in critical control	Certification limits; Critical control without learning; No opaque models in control; Limits on learning systems; Frozen critical behaviour
Theme 2: Dual role of AI: development tool and non-critical adviser	Learning for code generation; Advisory computer vision; Assistive role of automation; Offline learning loop; Selective traceability
Theme 3: Redundancy and fault isolation as primary safety net	Redundant reconfiguration; Command and monitor pattern; Diverse redundancy; Triple system voting; No cascading failures; Monitor shutdown and takeover; Voting conflict resolution; Fault isolation
Theme 4: Human-centred authority and pilot engagement	Pilot final authority; Manual override options; Design contrast with Boeing; Flight phase roles; Preserve manual skills; Action focused feedback
Theme 5: Traceability, learning from failures and specification updates	Traceability for anomalies; Post flight log analysis; Specification gaps; Logs for design updates; Selective traceability

D

Second Focus Group

D.1. Focus Group Analysis

Open Coding

Table D.1: Open Coding Analysis – Second Focus Group (FG2)

Code Name	Definition	Source Extract
Prompting vs weight training distinction	Distinguishes model weight training/fine-tuning from prompting and orchestration; reframes the work as application-layer system design.	“Do you mean re-training or fine-tuning model weights, or prompting... and building a system around it?”
Training is expensive / not the focus	Positions re-training as costly and not the intended approach; emphasizes evaluation and design around a largely fixed base model.	“Re-training model weights is expensive and technically different.”
Disruption readiness as evaluation, not retraining	Defines disruption readiness as validation of system behaviour under rare events rather than improving generic training metrics.	“The key principles are about system evaluation and testing, not re-training.”
GenAI/Agentic AI not ideal for detection	Questions the suitability of LLM-based agents as primary mechanisms for detecting out-of-distribution disruptions.	“...whether GenAI or Agentic AI is the best tool for detecting out-of-distribution events.”

Continued on next page.

Table D.1 continued.

Code Name	Definition	Source Extract
Specialised anomaly detection preferred	Frames anomaly detection and forecasting as better handled by specialised models or classical approaches than by agentic LLM systems.	"Detection is usually better handled by specialised approaches (e.g., anomaly detection and forecasting methods)."
Two-step architecture: detect then respond	Separates detection from response: specialised detectors trigger agentic workflows for mitigation and execution.	"...a two-step approach: detect first, respond second."
Agents lack native awareness of distributions	States that LLM agents do not inherently know what is "normal" vs. out-of-distribution without explicit statistical signals.	"The agent does not automatically 'know' what is normal versus out of distribution..."
Use explicit detection signals (scores/thresholds)	Operationalises disruption detection via structured indicators (anomaly scores, thresholds, confidence measures) integrated into the agent system.	"...provide signals such as anomaly scores, thresholds, confidence measures..."
Scenario-based stress testing	Defines robustness testing as constructing disruptive scenarios and evaluating behaviour under those stress conditions.	"You synthetically create disruptive scenarios and test the system's behavior."
Synthetic scenarios may be unrealistic	Warns that synthetic scenario generation can produce implausible or internally inconsistent cases, undermining validity.	"...scenarios that are not realistic or internally consistent."
Define scenario families (boundary/extreme/-plausible)	Recommends structured scenario families (extreme-but-plausible, boundary, intermediate) to systematically probe crisis behaviour.	"Define scenario families—extreme-but-plausible cases, boundary cases, and intermediate cases."
Predefined desired outcomes	Stresses specifying expected behaviours and acceptable outcomes for edge cases instead of relying on generic performance metrics.	"...specify expected response characteristics."
Action boundaries under disruption	Requires explicit constraints on permissible actions during disruption states to prevent unsafe or uncontrolled execution.	"Which actions are allowed under a disruption state..."
Escalation vs autonomy rules	Highlights the need to define when the agent may act autonomously versus when it must escalate to humans.	"...handle autonomously versus escalate..."

Continued on next page.

Table D.1 continued.

Code Name	Definition	Source Extract
Unit + end-to-end testing mindset	Applies software-testing logic to agents: component tests plus full workflow tests for crisis conditions.	"Unit tests and end-to-end tests for agents."
Workflow testing chain	Specifies testing across the full action sequence, including recovery mechanisms, to prevent brittle end-to-end behaviour.	"Detection signal → diagnosis → action proposal → execution → monitoring → rollback/escalation..."
Cascading error risk	Notes that multi-step agent action can amplify a single error into downstream operational harm.	"...one incorrect step can propagate downstream."
Synthetic data generalisation risk	Identifies sim-to-real and representativeness as core limitations: synthetic data may not transfer to real disruptions.	"Synthetic data may not generalize to real disruptions."
Simulation realism is difficult	Recognises that high-fidelity simulation of disruption dynamics is difficult and resource-intensive as complexity increases.	"Creating realistic simulations of disruption dynamics is difficult..."
Scoping makes scenario space manageable	Argues that focusing on bounded segments/disruption modes makes both scenario generation and evaluation tractable.	"You focus on a specific supply chain... When you scope it properly, the scenario space becomes manageable."
Probability cut-off / relevance boundary	Emphasises defining a boundary for rare events to avoid unbounded scenario spaces while still covering meaningful risk.	"You need a probability cut-off or relevance boundary."
Simulation benefit: scale and coverage	Highlights that synthetic scenarios enable broad coverage across multiple disruption types for stress testing.	"Scale and coverage... stress test across a broad range of conditions..."
Test under degraded observability	Recommends testing under poor data quality and partial visibility, reflecting real disruption conditions.	"...testing behavior under degraded data quality or partial observability."
Acceptable behaviour criteria	Defines success in behavioural terms: escalation, safety adherence, consistency, rollback, and limiting cascading harm.	"Correct escalation... adherence to action boundaries... recoverability (including rollbacks)..."
Agents more useful post-disruption	Positions agentic systems as most valuable after a disruption, supporting choice of actions in a changed operational reality.	"...most useful after a disruption occurs... decide the next best actions..."

Continued on next page.

Table D.1 continued.

Code Name	Definition	Source Extract
Shadow deployment as risk mitigation	Advocates running agents in parallel without execution to compare outputs with humans/legacy systems before deployment.	"Use shadow deployment: run the system in parallel... before giving it any operational control."
Start small and bounded	Reinforces that agent value depends on tight operational scoping and clearly defined tasks.	"Start very small—on a specific task."
Precision over generic problem framing	Warns that broad, generic tasks yield generic outputs; precise framing enables reliable and evaluable agent behaviour.	"...broad, generic problems... you tend to get generic answers. Precision... matters."
Comparative advantage of GenAI: unstructured text	Recommends using GenAI/agents where they outperform alternatives: unstructured text, synthesis, reasoning, and orchestration.	"...strongest is unstructured text, reasoning, and orchestration."
Regulation/external constraints as best-fit use case	Identifies external constraint monitoring as feasible and impactful due to its reliance on text comprehension and synthesis.	"Regulation and external constraint monitoring... strong fit for GenAI."
Need structured supply chain representation	States that text monitoring becomes actionable only when linked to a structured supply chain model for exposure mapping.	"If connected to a structured representation of the supply chain... reasoning about impacts..."
Early anomaly recognition to prevent amplification	Frames early detection as high leverage because upstream disruptions can amplify downstream (Bullwhip Effect).	"...disruptions early in the chain can amplify downstream (bullwhip effects)."
Transparency limits beyond Tier 1/2	Notes that limited multi-tier visibility constrains what any AI system can achieve for resilience.	"...often not available beyond Tier 1 or Tier 2 suppliers."
Supplier capability mapping as resilience enabler	Supports capability mapping as resilience-relevant, especially when supplier information is dispersed across documents and teams.	"Supplier capability mapping... GenAI helpful for reading and structuring supplier information."
Insights agent must be scoped	Accepts "insights" only when defined as disruption diagnosis and options generation, not as vague reporting.	"Insights / decision support... valuable if scoped... explaining why issues occur and recommending structured playbooks."

Continued on next page.

Table D.1 continued.

Code Name	Definition	Source Extract
Portfolio of bounded agents (shortlist)	Validates a portfolio approach (multiple bounded agents) rather than one generic resilience agent.	"Confirm the top five?... Demand... Regulation... Insights... Supplier capability mapping..."

Thematic Analysis

Table D.2: Themes and Supporting Open Coding Names – Second Focus Group

Theme Name	Associated Codes
Theme 1: Build readiness into the system (not retraining)	Prompting vs weight training distinction; Training is expensive / not the focus; Disruption readiness as evaluation, not retraining
Theme 2: Detect first, then respond	GenAI/Agentic AI not ideal for detection; Specialised anomaly detection preferred; Two-step architecture: detect then respond; Agents lack native awareness of distributions; Use explicit detection signals (scores/thresholds)
Theme 3: Stress-test with scenarios and clear expectations	Scenario-based stress testing; Synthetic scenarios may be unrealistic; Define scenario families (boundary/extreme/plausible); Predefined desired outcomes; Probability cut-off / relevance boundary; Simulation benefit: scale and coverage; Simulation realism is difficult; Synthetic data generalisation risk; Test under degraded observability
Theme 4: Guardrails, escalation, and recovery	Action boundaries under disruption; Escalation vs autonomy rules; Unit + end-to-end testing mindset; Workflow testing chain; Cascading error risk; Acceptable behaviour criteria
Theme 5: Start in shadow mode and keep scope tight	Shadow deployment as risk mitigation; Start small and bounded; Precision over generic problem framing; Agents more useful post-disruption
Theme 6: Focus on best-fit, high-impact use cases	Comparative advantage of GenAI: unstructured text; Need structured supply chain representation; Regulation/external constraints as best-fit use case; Supplier capability mapping as resilience enabler; Insights agent must be scoped; Early anomaly recognition to prevent amplification; Transparency limits beyond Tier 1/2; Portfolio of bounded agents (shortlist)

Table D.3: Themes and Definitions – Second Focus Group

Theme Name	Definition
Theme 1: Build readiness into the system (not retraining)	Preparation for rare events is accomplished by structuring the surrounding socio-technical system—including prompts, orchestration, controls, and evaluation—around a largely fixed base model, rather than pursuing expensive model retraining.

Continued on next page.

Table D.3 continued.

Theme Name	Definition
Theme 2: Detect first, then respond	Effective disruption management necessitates a bifurcated architecture in which specialized models detect anomalies and generate structured signals, while agentic workflows address diagnosis, coordination, and execution of mitigation measures following disruption identification.
Theme 3: Stress-test with scenarios and clear expectations	Effective disruption readiness requires systematic scenario generation and rigorous stress testing. This process involves defining scenario families, establishing boundaries, and specifying expected outcomes. Additionally, testing should be conducted under edge conditions, such as degraded data, while actively addressing simulation-to-reality and representativeness risks.
Theme 4: Guardrails, escalation, and recovery	Preventing cascading failures requires governing agent autonomy with explicit action constraints, escalation rules, and recoverability mechanisms, all of which should be validated through component and end-to-end testing across complete workflows.
Theme 5: Start in shadow mode and keep scope tight	Operational deployment should initially occur in shadow mode and focus on narrowly defined tasks. Agents provide the greatest value by supporting decision-making and execution following disruptions and under new constraints, rather than by attempting broad, global resilience analysis.
Theme 6: Focus on best-fit, high-impact use cases	GenAI and agents should be deployed in contexts where they demonstrate superior performance compared to alternative approaches, such as text comprehension, synthesis, reasoning, and orchestration. Their application should focus on high-impact resilience use cases, including external constraints monitoring, supplier capability mapping, scoped diagnosis and options support, and early anomaly interpretation. It is also necessary to recognize transparency limitations beyond Tier 1 and Tier 2 and to implement a portfolio of bounded agents.

D.2. First Follow-Up Interview

Open Coding

Table D.4: Open Coding Analysis - First Follow-Up Interview

Code Name	Definition	Source Extract
Legacy vs New-Build Implementation	Implementation mode depends on whether a client has an existing system to improve (legacy environment) versus starting from scratch (new-build).	“There are two options. Either the client already has something we need to improve … or we can start Greenfield.”
Reusable Asset Repository	A growing internal repository of agents, LLM components, and agentic networks reused across client engagements.	“We reuse all the assets we already have … in one large, growing repository that we bring to the client.”

Continues on next page.

Table D.4 continued.

Code Name	Definition	Source Extract
Client Constraints Drive Design	When clients already have systems, the solution must adapt to client-owned tools, data, and architecture rather than replacing them.	“Then we have to use what they have.”
Legacy Complexity as Opportunity	Suboptimal or complex client setups increase implementation effort (and revenue), even if frustrating for practitioners.	“The more suboptimal and complex it is, the better it is for us, even if it is frustrating.”
Minimal LLM Fine-Tuning	Fine-tuning or changing the base LLM is rare; standard components cover most use cases.	“Only once did I need to touch the large language model … All other times, we used pure standard components.”
RAG as Default Pattern	Retrieval-Augmented Generation is used routinely to ground outputs in internal documents and data.	“Yes, we use that all the time.”
Offline / Local Model Constraint	Running models locally (no internet / no external model calls) is a key risk for fit-for-purpose performance and deployment complexity.	“When we aren’t allowed to touch the internet … we need to run it locally, which can get tricky.”
Tool Connectivity Over Model Changes	Performance improvements come primarily from connecting the agent to relevant internal data and documents, not from retraining the LLM.	“We connect to the right internal data and documents.”
Out-of-Distribution Hallucination	When confronted with situations outside training coverage, current tools tend to fabricate plausible but false content.	“If it is truly not in the training data, they will come up with something else.”
Guardrails as Core Delivery	Safety controls are implemented as part of deployment to constrain outputs and prevent unacceptable behavior.	“We ensure there are guards at the end.”
Learned Outcome Validator Agent	A secondary agent can be trained to judge whether outputs are acceptable (meta-evaluation / critique layer).	“We might teach another agent what acceptable outcomes look like.”
Hard Constraints for Numeric Outputs	Rule-based boundaries enforce outputs within known valid ranges (especially for numbers).	“Use hard guards to predict numbers within specific known ranges.”

Continues on next page.

Table D.4 continued.

Code Name	Definition	Source Extract
Neuro-symbolic Control	Hybrid approach: neural models generate or infer, while symbolic rules constrain and validate for the specific use case.	“Neuro-symbolic AI … merges … language models … with symbolic logic (rules engines).”
Disruption Identification Problem	A key challenge is distinguishing a real disruption (novel reality) from model error; detecting “what counts” as disruption is hard.	“The hardest problem is determining when something is a disruption.”
Text Robustness Unresolved	Robust handling of disruptions remains weaker for text summarisation than for numeric prediction.	“For text summaries, it is still an open question.”
Human-in-the-Loop Default	Humans remain embedded in workflows due to uncertainty, novelty, and unresolved robustness issues.	“That is why we still have a human in the loop everywhere.”
Scenario Enumeration Mindset	Practitioners aim to proactively list most plausible failure scenarios based on client understanding.	“We can think of 90% of the scenarios that can go wrong.”
Synthetic Event Generation	Unexpected events are created synthetically so the system learns response patterns for events not seen historically.	“I then create those events synthetically to train the AI.”
Rule-based Simulation Training Data	Synthetic datasets are produced using “old-fashioned” rule-based simulations to create labeled patterns for learning or calibration.	“I simulate all these events using old-fashioned simulations.”
Severity Sensitivity Learning	Simulation varies disruption severity to teach the model thresholds and action relevance.	“I simulate a train being one minute late versus 40 minutes late.”
Threshold-based Action Policy	The system learns decision thresholds (e.g., when delay becomes problematic) even if the threshold event never happened historically.	“More than 40 minutes is a problem while less is fine.”
Digital Twin Cost-Benefit Tradeoff	Digital twins may be less attractive than expanding training data because building/maintaining a twin can be costly.	“It is almost more work to create a digital twin now than it is to just feed the model more data.”

Continues on next page.

Table D.4 continued.

Code Name	Definition	Source Extract
Digital Twin Fragility Under Uncertainty	Digital twins degrade when uncertainty or human behavior becomes significant, reducing reliability for complex socio-technical systems.	“As soon as there is uncertainty or a human involved, a digital twin tends to break down.”
Criticality-based Escalation	Decision to stop automation and call a human depends on task criticality and acceptable failure consequences.	“That depends on the importance of the task.”
Fail-and-Reset Tolerance	Non-critical tasks may be allowed to fail and be retried or reset without escalation.	“Some tasks can fail and just be reset.”
Stop-on-Doubt Policy	For critical tasks, any uncertainty triggers human intervention.	“If there is even the slightest doubt, we stop the process and call a human.”
Redundant Agent Consensus	Multiple agents perform the same task; disagreement acts as an uncertainty signal that triggers escalation.	“Four agents do the same thing; if one gives a different result, the human is called.”
Retry-until-Valid Loop	Automation may iterate until outputs meet strict rules when the task is structured and rule-checkable.	“An agent can try again until the end state adheres to the rules.”
Human Shadow Mode	Humans and agents operate in parallel to benchmark agent performance against human work.	“A human and an agent doing the same work in parallel.”
Shadow Agentic Networks	A second independent agentic workflow runs in parallel to monitor reliability and drift relative to the primary workflow.	“A separate agentic network does the same thing as the main one.”
Non-determinism Monitoring	Because agent outputs vary, monitoring focuses on divergence over time as a signal of instability or drift.	“These agents are not deterministic … we monitor if they start to deviate.”
Robustness as Recovery Speed	Robustness is operationalised as the speed of adaptation after a novel event, rather than perfect performance at onset.	“The key metric is how fast the models adapt to reality after an unexpected event.”
Rapid Detection and Rerouting	Value in disruptions comes from quickly detecting the new reality and adjusting flows (e.g., rerouting volumes).	“Quickly detecting that volumes must now move to other specific ports.”

Continues on next page.

Table D.4 continued.

Code Name	Definition	Source Extract
Continuous Exposure to Diverse Events	Ongoing learning requires continuous feeding with varied events, not one-time training.	“Constantly feeding it with a variety of events.”
Root Cause Relationship Learning	Teach causal links between disruptions and upstream signals so the system can reason beyond symptoms.	“Teach the model the relationship between events through root cause analysis.”
Proactive Signal-based Anticipation	Use leading indicators (e.g., weather, commodity prices) to anticipate disruptions and act proactively.	“Teaching the model to look at the real signal … allows it to handle the situation perfectly and even become proactive.”

Table D.5: Themes and Supporting Open Coding Names – Follow-Up Interview (1st Participant)

Theme Name	Supporting Open Coding Name
Theme 1: Implementation Architecture and Constraints	Legacy vs New-Build Implementation; Client Constraints Drive Design; Legacy Complexity as Opportunity; Minimal LLM Fine-Tuning; Reusable Asset Repository
Theme 2: Grounding and Output Control	RAG as Default Pattern; Tool Connectivity Over Model Changes; Offline / Local Model Constraint; Out-of-Distribution Hallucination; Guardrails as Core Delivery; Hard Constraints for Numeric Outputs; Learned Outcome Validator Agent; Neuro-symbolic Control
Theme 3: Disruption Management and Detection	Disruption Identification Problem; Text Robustness Unresolved; Human-in-the-Loop Default; Scenario Enumeration Mindset; Synthetic Event Generation; Rule-based Simulation Training Data; Severity Sensitivity Learning; Threshold-based Action Policy
Theme 4: Risk-Based Autonomy Architecture	Criticality-based Escalation; Fail-and-Reset Tolerance; Stop-on-Doubt Policy; Redundant Agent Consensus; Retry-until-Valid Loop; Non-determinism Monitoring
Theme 5: Validation and Monitoring Strategies	Human Shadow Mode; Shadow Agentic Networks; Non-determinism Monitoring; Digital Twin Cost-Benefit Tradeoff; Digital Twin Fragility Under Uncertainty
Theme 6: Operational Robustness as Adaptive Capability	Robustness as Recovery Speed; Rapid Detection and Re-routing; Continuous Exposure to Diverse Events; Root Cause Relationship Learning; Proactive Signal-based Anticipation

Thematic Analysis

Table D.6: Themes and Definitions – First Follow-Up Interview

Theme Name	Definition
Theme 1: Implementation Architecture and Constraints	AI implementation is shaped by client starting conditions (legacy enhancement vs. greenfield builds) and emphasizes standardized, reusable components over base model modification, with teams maintaining asset repositories and adapting to pre-existing infrastructure constraints.
Theme 2: Grounding and Output Control	Systems achieve reliability through two complementary mechanisms: RAG and tool connectivity for contextual grounding in enterprise data, and guardrails (hard constraints, outcome validators, neuro-symbolic control) that prevent out-of-distribution hallucinations and bound outputs within acceptable ranges.
Theme 3: Disruption Management and Detection	Disruptions are treated as novel realities requiring distinct handling: detection mechanisms distinguish genuine environmental shifts from model errors, while synthetic scenario generation and rule-based simulations proactively prepare systems for plausible failure modes and severity variations absent from historical data.
Theme 4: Risk-Based Autonomy Architecture	Autonomy boundaries are determined by task criticality: non-critical work tolerates fail-and-reset cycles or retry-until-valid loops, while critical tasks implement stop-on-doubt escalation policies, with redundant agent consensus mechanisms surfacing uncertainty through output divergence.
Theme 5: Validation and Monitoring Strategies	Non-deterministic agent behavior requires continuous validation through parallel execution (human shadow modes or shadow Agentic networks) that monitor drift and divergence; digital twins offer an alternative but face cost-benefit challenges and fragility under uncertainty, making them less practical than data-driven approaches.
Theme 6: Operational Robustness as Adaptive Capability	Robustness is operationalized as rapid post-disruption adaptation speed rather than initial perfect performance, achieved through continuous exposure to diverse events, explicit learning of root cause relationships, and development of proactive anticipation capabilities using leading indicators and upstream signals.

D.3. Second Follow-Up Interview

Open Coding

Table D.7: Open Coding Analysis - Second Follow-Up Interview

Code Name	Definition	Source Extract
Opaque training data	Users cannot know what the model has seen; OOD is inferred indirectly from behaviour.	“We usually don’t know exactly what’s in the training data of these models.”

Continues on next page.

Table D.7 continued.

Code Name	Definition	Source Extract
OOD failure pattern: over-complex outputs	When faced with unfamiliar tasks, tools may produce verbose, convoluted, or nonsensical solutions rather than concise ones.	“If you’re doing something out of the ordinary, these tools often don’t manage to implement a nice, concise solution... the solution becomes overly complex and not really to the point.”
Coding tools reveal limits fastest	AI limitations become most visible in code generation contexts compared to plain text interaction.	“You notice the limitations most quickly in AI coding tools.”
Hallucination on private specifics	The system invents details when asked about user/project information it cannot access.	“It starts hallucinating when I talk about my own project details.”
Privacy disclaimer behaviour	Newer models hedge or disclaim lack of access to private information when prompted about it.	“With the latest models it usually gives some kind of disclaimer that it isn’t aware of private information.”
Knowledge cut-off limitation	Pure GenAI becomes unreliable on facts/events beyond its training cut-off unless external retrieval is used.	“It fails when you ask about things beyond its knowledge cut-off point.”
Search tool mitigation	Search/browsing features can partially reduce cut-off failures for current events.	“Most tools nowadays have search functionality, so they can find information about current affairs.”
Out-of-scope / private-data boundary	Models fail or become unreliable when asked for information they objectively cannot have (private, restricted, out of scope).	“It’s mostly when... the model doesn’t have the information—like when you ask about private data or something out of scope.”
Context window degradation	Near maximum context length, performance drops and the system loses track of requirements/state.	“When you’re close to maximising the context window, it can become unreliable... they start losing track.”
Training = configuration (not weights)	In enterprise deployments, “training” often means configuring workflows, tools, and instructions rather than retraining weights.	“When you say ‘training,’ you mean configuring it... not necessarily retraining model weights.”
No weight retraining in client work	Typical implementation avoids retraining foundation models; emphasis is on process mapping and orchestration.	“You’re not going to retrain the underlying models.”
Agentic AI as automation mapping	Agentic AI is framed as process automation: translate a business process into an agent network.	“Agentic AI, in principle, is an automation problem... map the specific business process you want to automate.”

Continues on next page.

Table D.7 continued.

Code Name	Definition	Source Extract
Task decomposition into agents	Break work into subtasks, define specialised agents, and assign targeted instructions.	“For specific subtasks, you define an agent and give it a very specific set of instructions.”
Tool permissioning	Agents are configured by specifying allowable tools (e.g., web search, SQL), constraining capabilities.	“You specify the tools it can use—web search, running SQL queries, and so on.”
Agent communication structure	Configuration includes which agents can communicate and how coordination occurs.	“You also specify which other agents it can communicate with.”
Orchestrated execution order	Many agentic systems rely on predefined sequencing (plan first, then execute), not free-form autonomy.	“You already know the order in which you want your agents to function... first... a research plan... then... a web-search agent.”
Specific-task configuration outperforms general autonomy	Narrowly scoped agents tend to be more reliable than general autonomous agents because instructions and boundaries are clearer.	“When you configure agents for a more specific task, it usually works better than a general autonomous agent.”
Rule-like handling guidance	Use explicit conditional instructions (“if X then Y”) to steer behaviour in anticipated situations.	“You can give good directions and hints—like: ‘If you encounter situation X, do Y.’”
General autonomy uncertainty gap	Truly general agents are harder to control because the state space of possible situations is unknown.	“That’s difficult with a more general, autonomous agent, because you don’t know what situations it will run into.”
Synthetic data as necessity under scarcity	When rare-event history is unavailable, synthetic/simulated scenarios become essential for configuration and evaluation.	“If you have no historical data available, then your only solution is synthetic data or simulation-based approaches.”
Train–test split for agent configuration	Use held-out scenarios to test generalisation and avoid overtuning to known cases.	“Use one subset to configure... and another subset... to test whether your instructions generalise.”
Overfitting to scenario set	Using all scenarios for configuration risks tuning too specifically to those cases, reducing robustness.	“If you use all your data to configure the agents, they might become too specifically tuned to those situations.”
Configuration is “hope it works”	Agent setup is not guaranteed; configuration is iterative and outcome depends on use case quality and constraints.	“You’re not actually training a model—you’re configuring something and hoping it works.”

Continues on next page.

Table D.7 continued.

Code Name	Definition	Source Extract
Domain expert validation loop	Domain experts are needed to validate scenario realism and completeness because builders lack domain expertise.	“Show them your examples... ‘Am I missing something? Is this realistic?’... you usually aren’t the domain expert.”
Benefit: quality improvement	Primary value pathway is reducing human errors and improving output quality.	“Either quality improves (fewer human errors)...”
Benefit: efficiency improvement	Primary value pathway is faster task execution and throughput.	“...or efficiency improves (it becomes faster).”
Consistency via well-defined agents	Well-specified agents can standardise execution, improving consistency (conditional on autonomy constraints).	“If you use well-defined agents, tasks are handled more consistently.”
Autonomy erodes consistency	High autonomy reduces predictability; the agent may deviate from desired consistent routines.	“If it’s entirely autonomous, that consistency can fade, because it can still do whatever.”
Synthetic data benefit: enables evaluation	Synthetic data provides a minimum viable dataset to run tests when real data is missing.	“If your options are ‘no data’ or ‘synthetic data,’ at least you have something to work with... run evaluations and test multiple scenarios.”
Synthetic data risk: non-representative scenarios	If synthetic scenarios don’t match reality, evaluation results mislead and add little practical value.	“The main risk is that the synthetic data isn’t representative of real-life scenarios.”
Synthetic data risk mitigation: realism checks	Use expert judgement to flag unrealistic scenarios and improve representativeness.	“A domain expert might say, ‘This is nonsensical’ or ‘This is not realistic.’”
OOD recognition importance depends on impact	OOD detection is critical in high-impact/no-review contexts; less critical when outputs are checked and consequences are low.	“In high-impact situations with no evaluation... it’s really important... If it’s a safer environment where people check outputs anyway, I’d let it try.”
RAG similarity threshold for OOD flag	Implement objective OOD detection by thresholding retrieval similarity in a RAG pipeline.	“With RAG, you can set a threshold on retrieval similarity... this is out-of-distribution.”
Prompt-defined action boundaries	Alternative is to encode action/no-action boundaries in prompts, but this is harder and designer-dependent.	“Careful prompting... explicitly explaining when it should act and when it should not... but that’s harder.”
Prompt designer bias	Behaviour boundaries reflect the creator’s judgement; prompting introduces subjective bias.	“Absolutely. It’s up to the creator and the person writing the prompts.”
Default response: flag unexpected + explain	Minimum safe default is to signal unexpectedness and provide reasoning/justification.	“At minimum, I would let it indicate that it detected something unexpected... perhaps... explain its reasoning.”

Continues on next page.

Table D.7 continued.

Code Name	Definition	Source Extract
High-impact default: refuse / do not answer	In customer-facing or high-stakes contexts, the safest default is refusal rather than guessing.	“In high-impact situations, you may want it to default to not answering... ‘Sorry, I can’t answer this question.’”
Logging for continuous improvement	Unexpected cases should be logged with rationale and reused as evaluation cases to improve future performance.	“Log the specific scenario and why it judged it unexpected... include it in the evaluation set next time.”
Robustness evaluation via KPI comparison	Assess robustness by comparing KPI performance on in-scope versus unexpected scenarios.	“Define measurable KPIs... see whether the model performs equally well... on unexpected situations.”
Objective metrics when ground truth exists	When outputs have correctness criteria (e.g., code compiles), use objective accuracy-like measures.	“For coding tasks: does the code compile, and is the output correct?”
Rubric + LLM-as-judge for fuzzy tasks	For non-binary tasks, score outputs with a rubric and automated judging to approximate evaluation at scale.	“Define a rubric... then use an LLM-as-a-judge... to automatically score outputs.”
Priority: detect unexpected events	Top improvement is equipping systems to detect anomalies relative to normal baselines/patterns.	“Give it tools and options to detect that something is unexpected... aware of regular demand patterns.”
Priority: clear handling instructions	Explicit guidance on what to do in unexpected situations is essential to reduce unsafe improvisation.	“You need clear instructions on how you want it to handle unexpected situations.”
Two-step architecture: detect then respond	Separate detection from response: anomaly detection identifies the event; the agent supports interpretation and action planning.	“One component detects whether it’s an unexpected event, and another responds.”
Use deterministic methods when reliable	Prefer deterministic/statistical tools when they work; use GenAI when data is unstructured or ambiguity remains.	“If there is a deterministic tool that works reliably, I would always use the deterministic tool... deterministic models require structured data... if unstructured... resort to generative methods.”
Complexity minimisation principle	Start with the simplest approach; increase complexity only if needed because multi-agent systems are difficult to maintain.	“Start with the least complex tool... a multi-agent system is on the higher end of complexity... maintaining it becomes painful.”
Ongoing inference cost consideration	Agentic systems carry recurring operational costs (inference), unlike many deterministic models.	“With agents, you usually have ongoing inference costs... costs keep accumulating during operations.”

Thematic Analysis

Table D.8: Themes and Supporting Open Coding Names – Second Follow-Up Interview

Theme Name	Supporting Open Coding Name
Theme 1: Practical OOD failure modes and system limits	Opaque training data; Coding tools reveal limits fastest; OOD failure pattern: overcomplex outputs; Hallucination on private specifics; Privacy disclaimer behaviour; Knowledge cut-off limitation; Context window degradation; Out-of-scope / private-data boundary
Theme 2: Enterprise "training" as configuration and orchestration	Training = configuration (not weights); No weight retraining in client work; Agentic AI as automation mapping; Task decomposition into agents; Tool permissioning; Agent communication structure; Orchestrated execution order
Theme 3: Disruption handling via constrained autonomy	Specific-task configuration outperforms general autonomy; Rule-like handling guidance; General autonomy uncertainty gap; Autonomy erodes consistency; Consistency via well-defined agents
Theme 4: Scenario-based engineering and expert-grounded realism	Synthetic data as necessity under scarcity; Train–test split for agent configuration; Overfitting to scenario set; Configuration is "hope it works"; Domain expert validation loop; Synthetic data risk: non-representative scenarios; Synthetic data risk mitigation: realism checks
Theme 5: Safety gates and safe-default recovery loops	OOD recognition importance depends on impact; RAG similarity threshold for OOD flag; Prompt-defined action boundaries; Prompt designer bias; Default response: flag unexpected + explain; High-impact default: refuse / do not answer; Logging for continuous improvement
Theme 6: Robustness evaluation and engineering priorities under cost constraints	Robustness evaluation via KPI comparison; Objective metrics when ground truth exists; Rubric + LLM-as-judge for fuzzy tasks; Priority: detect unexpected events; Priority: clear handling instructions; Two-step architecture: detect then respond; Use deterministic methods when reliable; Complexity minimisation principle; Ongoing inference cost consideration

Table D.9: Themes and Definitions – Follow-Up Interview (2nd Participant

Theme Name	Definition
Theme 1: Practical OOD failure modes and system limits	Out-of-distribution behavior manifests through observable failure patterns rather than known training boundaries: systems produce over-complex or nonsensical outputs when facing unfamiliar tasks (most evident in coding contexts), hallucinate details about private or out-of-scope information (though newer models increasingly provide privacy disclaimers), fail on post-cutoff knowledge without retrieval augmentation, and experience performance degradation near maximum context window utilization.
Theme 2: Enterprise "training" as configuration and orchestration	Enterprise AI implementation reframes "training" as system configuration rather than weight modification: practitioners avoid foundation model retraining, instead treating Agentic AI as an automation mapping problem where business processes are decomposed into specialized agent networks with explicit tool permissions, communication structures, and orchestrated execution sequences that prioritize predefined workflows over free-form autonomy.

Continued on next page.

Table D.9 continued.

Theme Name	Definition
Theme 3: Disruption handling via constrained autonomy	Robust unexpected-event handling emerges from narrow task scoping and explicit behavioral guidance: task-specific agent configurations enable rule-like conditional steering (”if X then Y”), predictable coordination, and improved consistency, while general autonomous agents face control challenges due to unknown situation spaces and the consistency erosion that accompanies broad autonomy.
Theme 4: Scenario-based engineering and expert-grounded realism	When historical data for rare events is unavailable, synthetic scenarios and rule-based simulations become essential for agent configuration and evaluation. Practitioners employ train-test splits on scenario sets to avoid overfitting to known cases, treating configuration as an iterative ”hope it works” process that relies critically on domain expert validation loops to assess scenario realism, identify coverage gaps, and mitigate the risk of non-representative synthetic data misleading evaluation outcomes.
Theme 5: Safety gates and safe-default recovery loops	Unexpected situations require implementable detection mechanisms and conservative fallback behaviors, with criticality determining response stringency: high-impact/no-review contexts demand OOD recognition through objective mechanisms (e.g., RAG similarity thresholds) and safe defaults that prioritize refusal over speculation, while lower-stakes environments can tolerate uncertainty flagging with explanations; all unexpected cases should be systematically logged with rationale for continuous improvement and future evaluation expansion.
Theme 6: Robustness evaluation and engineering priorities under cost constraints	Robustness is operationalized through KPI-based performance comparison between in-scope and unexpected scenarios, using objective metrics where ground truth exists (e.g., code compilation) and rubric-based LLM-as-judge evaluation for ambiguous tasks. Engineering priorities emphasize two-step architectures (detect then respond), preference for deterministic methods over GenAI when reliable, complexity minimization to reduce maintenance burden, and explicit consideration of ongoing inference costs that accumulate during Agentic system operations.