EISIC 28 - 2025

Rethinking Business Continuity: A Critical Interpretability Perspective on Human-AI Crisis Integration

Davide Liberato lo Conte

Sapienza University of Rome, Italy

davideliberato.loconte@uniroma1.it

Giuseppe Sancetta

Sapienza University of Rome, Italy

giuseppe.sancetta@uniroma1.it

Valerio Antonini

Dublin City University, Ireland

valerio.antonini3@mail.dcu.ie

Abstract

Purpose

As artificial intelligence (AI) systems become integral to crisis response and continuity planning, organizations face a fundamental challenge: how to preserve human judgment in machine-rich decision environments. This study investigates how interpretive capacity—the ability to question, contextualize, and ethically manage AI outputs—shapes business continuity under conditions of volatility, uncertainty, complexity, and ambiguity. It introduces *critical interpretability* as a cognitive–organizational capability that determines whether AI enhances or undermines resilience.

Methodology

Using a multiple case study design, the research examines ten Italian firms across high-impact sectors (energy, health, logistics, and finance) that deployed AI tools during crises between 2020 and 2024. Data were collected through interviews with executives, data scientists, and crisis managers, supplemented by internal documentation and archival materials. A grounded theory approach enabled the inductive development of a multi-level framework connecting human cognition, organizational design, and AI system features.

Findings

The results show that effective continuity arises not from technological sophistication alone, but from the depth of human engagement with AI. When decision-makers actively interpret, challenge, and ethically calibrate algorithmic recommendations, continuity responses become adaptive and contextually sound. The study identifies three enabling layers—AI system capabilities, human interpretive practices, and organizational conditions—whose alignment supports ethical and resilient action. Misalignment, by contrast, leads to brittle automation and ethical blind spots.

Research limitations and implications

While limited to a specific national and temporal context, the study provides a strong conceptual foundation for future cross-national and longitudinal research on interpretability and resilience. It suggests new directions for measuring interpretive maturity and for integrating cognitive and ethical dimensions into AI governance frameworks.

Originality and value

The paper challenges techno-deterministic perspectives by repositioning AI not as a replacement for human cognition, but as a catalyst for interpretive and ethical reasoning. It presents the first empirically grounded framework of *critical interpretability* in the context of organizational continuity, offering novel insights for scholars and practitioners seeking to balance automation, accountability, and adaptive sensemaking in high-stakes environments.

Keywords

Artificial Intelligence; Crisis Management; Business Continuity; Human–AI Collaboration; Critical Interpretability.

Paper type

Research paper.

1. Introduction

In an age increasingly defined by disruption—geopolitical, environmental, technological, and social—organizations face a profound managerial question: *what does it truly mean to remain resilient amid uncertainty?* From pandemics and energy crises to cyberattacks and supply chain collapses, contemporary disruptions are not isolated anomalies but interconnected patterns of volatility (Boin & van Eeten, 2013). The "*new normal*" is one of cascading complexity, where uncertainty itself becomes a structural condition of organizational life (Benbya et al., 2020). Within this landscape**,** business continuity has evolved beyond its traditional function as a technical or procedural safeguard. It now represents a strategic and ethical imperative—a dynamic process through which organizations sustain not merely operations, but integrity, adaptability, and meaning. Continuity, in this sense, is not a checklist to be followed but a capability to be cultivated: a form of managerial cognition grounded in reflection, interpretation, and moral awareness. At the same time, Artificial Intelligence (AI) has become embedded in the architectures of decision-making and crisis response. Algorithms forecast disruptions, allocate resources, and recommend courses of action once entrusted to human expertise (Chatterjee et al., 2021; Dwivedi et al., 2021). The promise is compelling—speed, scalability, and data-driven precision—but it comes with a hidden cost: the gradual outsourcing of human judgment (Burton et al., 2020). As decision authority shifts from experienced professionals to opaque systems, organizations risk losing the very interpretive capacities that make resilience possible. This tension exposes what can be called the cognitive–ethical gap of AI-driven crisis management. While machines excel at identifying patterns, they cannot comprehend context, ambiguity, or value conflicts. Crises, however, are precisely defined by those features. They require not just faster analysis, but wiser sensemaking, decisions that unite information with judgment, foresight with ethics, and automation with accountability (Boin & van Eeten, 2013; Dignum, 2018). In this paper, we argue that the future of resilience depends on re-centering the human interpretive function within AI-mediated decision systems. We advance the concept of critical interpretability—the human capacity to understand, question, and ethically calibrate algorithmic insights—as a foundational element of organizational continuity. Rather than treating interpretability as a technical property of models (Doshi-Velez & Kim, 2017), we conceptualize it as a *cognitive and moral capability* through which managers translate algorithmic outputs into contextually grounded, ethically defensible actions (Dignum, 2018). The argument aligns with recent research suggesting that the adoption of AI requires not only technical readiness but also cognitive and ethical assimilation within organizations (Dwivedi et al., 2021). It responds to calls for a "*responsible AI*" paradigm—an approach that integrates human values into the design and governance of intelligent systems (Dignum, 2018)—and to growing recognition that the success of digital transformation depends on the alignment between technological systems and human sensemaking (Benbya et al., 2020). Accordingly, this research is guided by three interrelated questions:
first, how does human interpretive engagement shape the adaptability and ethical quality of AI-assisted decision-making during crises? second, what cognitive and organizational mechanisms enable the exercise of critical interpretability—allowing managers to question, reframe, and integrate algorithmic recommendations under uncertainty? and third, how can organizations design human–AI continuity systems that balance automation with accountability, ensuring that technological intelligence strengthens rather than supplants human judgment? By addressing these questions, this paper contributes to the emerging dialogue at the intersection of ethics, cognition, and technology. It positions resilience not as a product of predictive accuracy but as a *practice of interpretive responsibility*—the ability to act wisely, ethically, and reflectively in an increasingly automated world (Burton et al., 2020; Dignum, 2018; Doshi-Velez & Kim, 2017).

2. Literature Review

2.1. Business Continuity and Organizational Resilience in the Age of Disruption

Over the past two decades, business continuity has evolved from a narrowly technical concern to a strategic and cognitive capability—the organizational capacity to anticipate, interpret, and ethically navigate disruption. Traditional continuity planning focused on operational recovery; contemporary perspectives instead link continuity to learning, adaptation, and sensemaking (Boin & van Eeten, 2013; Hodgkinson & Healey, 2011). This shift reflects the reality that disruptions are no longer episodic anomalies but structural features of complex, interconnected environments (Benbya et al., 2020). Resilience, therefore, is increasingly conceptualized as a *dynamic capability* rooted in reflection, anticipation, and ethical judgment. Within volatile, uncertain, complex, and ambiguous (VUCA) contexts, organizations must integrate continuity into decision-making and culture, relying on interpretive cognition rather than static plans. Yet, as digital transformation accelerates, this task becomes intertwined with the growing influence of AI in managerial processes. AI-based systems now participate in forecasting, monitoring, and crisis coordination, reshaping how organizations perceive and respond to uncertainty (Faraj et al., 2020). These technologies extend human capability but also create new cognitive and ethical vulnerabilities, including automation bias and algorithmic opacity (Glikson & Woolley, 2020; Grundner & Neuhofer, 2021). As Dwivedi et al. (2021) argue, the challenge of AI assimilation lies not only in technical readiness but in *ethical awareness* and *organizational learning*. Recent debates thus converge on a socio-technical understanding of resilience, emphasizing that continuity emerges from the interaction between technological systems and human interpretive judgment. Doshi-Velez & Kim (2017) advocate for a rigorous science of interpretability, while Dignum (2018) calls for the design of *responsible AI* grounded in human values and moral accountability. Together, these insights highlight that resilience in the digital age depends less on algorithmic precision than on *ethical interpretability*—the ability of human actors to understand, question, and ethically align machine intelligence within complex decision environments. In this view, business continuity is no longer a technical safeguard but a form of responsible cognition: the capacity to sustain purposeful, value-conscious action through the joint intelligence of humans and machines.

2.2. Artificial Intelligence and Crisis Response: From Automation to Augmentation

AI technologies—ranging from machine learning and predictive analytics to autonomous agents—have become central to how organizations anticipate, interpret, and respond to crises. By enabling real-time data processing and complex scenario simulation, AI enhances situational awareness, responsiveness, and coordination during disruptive events (Tarafdar et al., 2019). In domains such as supply chain management, cybersecurity, and emergency response, AI systems are increasingly embedded in business continuity architectures, transforming how risk is sensed, communicated, and mitigated.

Yet this acceleration of automation introduces a new set of ethical and cognitive challenges. Scholars have warned that, while AI amplifies sensing capabilities, it can simultaneously erode reflective judgment and contextual reasoning (Raisch & Krakowski, 2021). The risk is not technological failure but cognitive overreliance—the tendency for human decision-makers to defer uncritically to algorithmic recommendations, especially under pressure or uncertainty (Power et al., 2021). In high-stakes settings, such as crisis response, this can lead to what Weick (1993) described as a *collapse of sensemaking*: the disintegration of interpretive processes when complexity exceeds the bounds of comprehension.

Indeed, research on human–AI teaming shows that crises amplify both the need for and the fragility of collaboration between humans and intelligent systems. While AI contributes speed and precision, humans provide ethical reasoning, situational empathy, and adaptive framing—the very qualities required when data are incomplete or ambiguous (Weick & Sutcliffe, 2015; Williams et al., 2017). The challenge, then, is to design socio-technical systems that sustain interpretive balance: leveraging automation for efficiency while preserving human oversight for moral and contextual calibration.

Recent developments in explainable and responsible AI echo this imperative. Rai (2020) calls for a transition from "*black box" to "glass box*" systems, where algorithmic logic becomes transparent and contestable. Similarly, Raisch & Krakowski (2021) highlight the automation–augmentation paradox,

arguing that organizational resilience depends not on replacing human cognition, but on amplifying it through human–AI complementarity. In this view, AI should serve as a cognitive catalyst rather than an autonomous decision-maker.

Ultimately, the integration of AI in crisis management requires more than technical readiness—it demands ethical attentiveness and interpretive awareness. When crises unfold, decisions must balance data-driven precision with human discernment, ensuring that automation strengthens, rather than suppresses, organizational sensemaking and moral accountability (Power et al., 2021; Weick, 1993; Weick & Sutcliffe, 2015).

2.3. Critical Interpretability as a Cognitive–Ethical Capability

An emerging but underexplored theme in AI-enabled crisis management concerns how humans interpret, question, and reshape machine-driven insights when decisions carry strategic and moral consequences. Building on the *cognitive turn* in crisis and organizational studies (Weick, 1993; Maitlis & Christianson, 2014), interpretability should not be viewed merely as a technical feature of algorithms, but as a *cognitive–social process* enacted by human agents. In this sense, critical interpretability refers to the human capability to critically assess, contextualize, and, when necessary, reframe or override AI-generated recommendations within high-pressure, uncertain environments. This conceptualization extends current debates on explainable and accountable AI (Kroll et al., 2016; Rai, 2020), shifting the focus from *how algorithms explain themselves* to *how humans make sense of them*. It emphasizes interpretation as an ongoing act of judgment, situated at the intersection of cognition, ethics, and socio-technical design. As Raisch & Krakowski (2021) argue, effective human–AI collaboration depends on hybrid decision architectures in which humans retain ultimate sensemaking authority. Decisions made under crisis conditions involve ethical trade-offs, strategic ambiguity, and reputational stakes that cannot be resolved through automated reasoning alone (Power et al., 2021). Critical interpretability thus operates as a managerial safeguard—ensuring that technological precision remains anchored in human prudence and responsibility. Empirical and conceptual studies have begun to identify conditions that enable or hinder this capability. Factors such as trust calibration, cognitive diversity, ethical literacy, and the interactivity of human–AI interfaces are found to be essential in fostering interpretive awareness (Grundner & Neuhofer, 2021; Papagiannidis & Bourlakis, 2023). Meanwhile, research on digital twins and intelligent infrastructures underscores the growing need for real-time interpretability in complex, data-intensive systems such as supply chains (Ivanov & Dolgui, 2020). Ethical AI design frameworks further highlight that transparency and accountability must operate across levels—individual, organizational, and systemic—if human judgment is to remain effective under pressure (Mikalef & Pappas, 2022). Ultimately, critical interpretability represents both a *psychological foundation* and an *organizational dynamic capability* (Hodgkinson & Healey, 2011). It allows decision-makers to sustain reflection amid disruption, aligning AI-enabled insight with human values, contextual reasoning, and moral accountability. In crisis management, this capability becomes indispensable—not as a technical add-on, but as the cognitive and ethical core of resilient, human-centered decision-making (Paschen & Ferreira, 2020; Power et al., 2021).

3. Research Objectives

Despite the widespread integration of AI across strategic and operational domains, the human conditions that enable effective collaboration between humans and intelligent systems during crises remain insufficiently understood. Existing research has largely centered on algorithmic performance and ethical principles, yet it has paid limited attention to the interpretive and cognitive work carried out by decision-makers navigating AI-mediated environments. This gap leaves unresolved questions: under what conditions does AI strengthen organizational continuity, and when might it weaken it? What mechanisms allow humans to critically engage with algorithmic insights rather than passively accept or reject them? Understanding these dynamics is essential to developing resilience in environments where automation increasingly mediates judgment, foresight, and action. The overarching goal of this study is to theorize the role of human interpretive judgment in AI-enabled continuity planning, particularly under conditions of high uncertainty and systemic disruption. Rather

than framing automation and human control as opposing forces, the study explores how hybrid intelligence systems—where algorithmic processing and human sensemaking operate in tandem—can sustain resilience and ethical decision-making. Specifically, this research pursues four interrelated objectives:

- *To examine how human actors engage with AI systems during crises, including how they interpret, question, reframe, or override algorithmic recommendations in continuity decision-making.*

- *To identify the cognitive, organizational, and technological conditions that enable or inhibit effective human–AI collaboration under pressure.*

- *To conceptualize and define "critical interpretability" as a distinct managerial capability essential for resilience in AI-mediated contexts.*

- *To develop a multi-level framework linking individual interpretive practices, organizational structures, and technological affordances to business continuity outcomes.*

These objectives stem from an urgent managerial and societal need. As crises grow in scale and complexity, and as AI systems gain decision autonomy, ensuring continuity depends not only on technological sophistication but also on the preservation and institutionalization of human judgment. Developing the capacity to interpret, question, and ethically align AI-driven insights is thus both a theoretical frontier and a practical necessity. In doing so, this study advances understanding at the intersection of crisis leadership, business continuity, socio-technical governance, and responsible AI, offering a framework for organizations to balance automation with accountability and to cultivate resilience grounded in human insight.

4. Methodology

4.1. Research Design

To explore how human decision-makers interact with AI systems during crisis management and business continuity processes, this study adopts a qualitative, multiple-case design. Such an approach is particularly suited to examining complex socio-technical dynamics that unfold in real-world settings, where meaning, cognition, and ethics are deeply contextual and cannot be reduced to measurable variables. The focus is on how interpretive judgment and technological mediation coexist and evolve under conditions of disruption. Recent research has emphasized that understanding human–AI interaction requires close attention to the processes of sensemaking and collaboration that emerge in high-pressure, uncertain contexts (Maitlis & Christianson, 2014; Power & Kazda, 2021). A qualitative multi-case strategy enables such processual understanding by capturing both variation and depth across organizations that differ in sector, AI maturity, and exposure to crises. This design also aligns with recent calls to investigate AI-enabled resilience through situated inquiry rather than abstract generalization (Papagiannidis & Bourlakis, 2023). Ten large and medium-sized Italian enterprises were selected using theoretical sampling to maximize diversity in industry context, technological sophistication, and crisis exposure. Italy provides a particularly fertile setting: as a major European manufacturing hub facing overlapping disruptions—ranging from pandemics to energy volatility—it offers rich empirical opportunities to observe hybrid intelligence in action, where human and machine decision processes intertwine. Methodologically, the study draws inspiration from emerging approaches in ethical AI and accountable design (Kroll et al., 2016; Mikalef & Pappas, 2022), treating AI not only as a technical system but as a cognitive and moral actor within organizational practice. Insights from digital resilience and intelligent infrastructures (Ivanov & Dolgui, 2020; Paschen & Ferreira, 2020) inform the analysis of how AI supports anticipation, coordination, and recovery in turbulent environments. By combining contextual richness with theoretical depth, the research design seeks to capture the lived experience of human–AI collaboration, highlighting the interpretive, ethical,

and organizational mechanisms through which continuity is achieved—or compromised—when technology meets crisis.

4.2. Case Selection and Profiles

The selected firms span five sectors: energy and utilities, manufacturing, logistics and supply chain, financial services, and healthcare. All had implemented AI-enabled tools or platforms to support business continuity or crisis response at the strategic or operational level. This configuration allowed us to explore sectoral variation in AI-human dynamics and to assess how organizational structures and crisis types influence interpretive practices.

4.3. Data Collection

Data were collected over a period of eight months through a triangulated approach combining:

- Semi-structured interviews;
- Archival documents and internal reports;
- Observation and digital trace analysis (in 3 cases): where permitted, we observed AI-supported continuity simulations or retrospectives (e.g., post-crisis review meetings, dashboard use sessions), and analyzed user logs showing human interaction with AI systems.

4.4. Data Analysis

We used an abductive coding strategy, allowing iterative movement between theory and empirical data. The analysis unfolded in three stages:

1. Open coding of interview transcripts and documents;
2. Axial coding;
3. Cross-case synthesis, comparing patterns across firms to build a mid-range theoretical framework linking human-AI interaction with continuity performance.

To ensure credibility and internal validity, we conducted member checks and shared case summaries with firms for feedback. Differences in interpretation were discussed and incorporated into the revised coding.

4.5. Methodological Rigor and Justification

This study prioritizes contextual realism and theoretical saturation over representativeness. By focusing on diverse Italian firms handling real disruptions, we uncover nuanced insights into how managerial cognition, institutional norms, and AI architecture interact in moments of crisis. The use of multiple sectors and triangulated data sources enhances external validity and analytical generalization. Our choice to focus on Italy was intentional: the country combines high exposure to systemic risk with heterogeneous AI adoption, making it ideal for observing interpretive gaps, governance frictions, and cognitive tensions that may remain hidden in more digitally mature ecosystems.

5. Findings

Our cross-case analysis revealed three major themes that characterize how organizations interact with AI systems during crisis-driven continuity planning: (1) Human Interpretive Agency, (2) Organizational Enablers and Frictions, and (3) Socio-technical Accountability and Governance Tensions. These dimensions form the backbone of what we conceptualize as "*critical interpretability"*, a strategic and cognitive capability essential to ensuring resilient outcomes in AI-mediated environments.

5.1. Human Interpretive Agency: Between Deference and Reframing

In all cases, AI systems played a pivotal role in enabling situational awareness, forecasting, and automation of routine responses. However, effective continuity was never a product of AI alone. Instead, continuity emerged when human actors actively engaged with AI outputs, questioning assumptions, contextualizing predictions, or even overriding suggestions when organizational, ethical, or contextual nuances demanded it. For instance, in Case C (healthcare), emergency response staff routinely modified AI-generated triage recommendations to account for real-time patient influxes and ethical considerations. In Case A (energy), operators noticed that AI systems underpredicted peak loads during politically driven demand surges, triggering manual overrides. Conversely, in Cases D and I, blind trust in AI dashboards during cybersecurity and logistics crises led to delayed human intervention and suboptimal recovery timelines. This spectrum, from *deferential adoption* to *critical reframing*, is shaped by both cognitive orientation (training, experience) and cultural expectations around authority and automation.

"*You have to know when the AI is wrong, and that means you need to think critically, not just follow*" (Manager, Case F).

5.2. Organizational Enablers and Frictions: Structures That Support or Inhibit Critical Engagement

The capacity for critical interpretability was significantly influenced by organizational factors. Firms with cross-functional continuity teams, embedded AI-literacy training, and structured escalation paths enabled more nuanced and context-aware engagement with AI outputs. For example, in Case E (banking), continuity staff were trained to "*challenge the model"* under stress scenarios, using counterfactuals and red-teaming exercises. In contrast, firms with rigid hierarchies or siloed technical teams (Cases D, I, and G) often reported interpretability breakdowns. Staff either lacked the confidence to question AI outputs or did not understand the underlying logic, especially under time pressure.

Key enablers identified include:

- AI transparency tools and explainability layers;
- Interdisciplinary crisis teams (IT + ops + strategy);
- Ethical escalation protocols;
- Real-time simulation training.

Barriers included:

- Lack of interpretive authority;
- Fear of accountability or blame when questioning AI.

5.3. Socio-technical Accountability and Governance Tensions

Across several firms, questions emerged around who is responsible when AI fails during crisis. While technical teams often *"owned*" the systems, decision-making authority remained with business units, creating ambiguity and risk-averse behaviors. In Case G (utilities), AI incorrectly projected flood impact zones, leading to under-deployment of personnel. The firm later discovered that the training dataset lacked recent topographical data. No single unit was held accountable, a common theme across sectors. To address this, Cases A, C, and F implemented dual validation procedures, where any high-impact AI decision required both system and human sign-off. Others, such as Case H, instituted AI ethics boards to review high-risk models pre-deployment. These practices reflect a broader shift: from designing AI *for* autonomy to embedding AI *within* human-centered governance ecosystems.

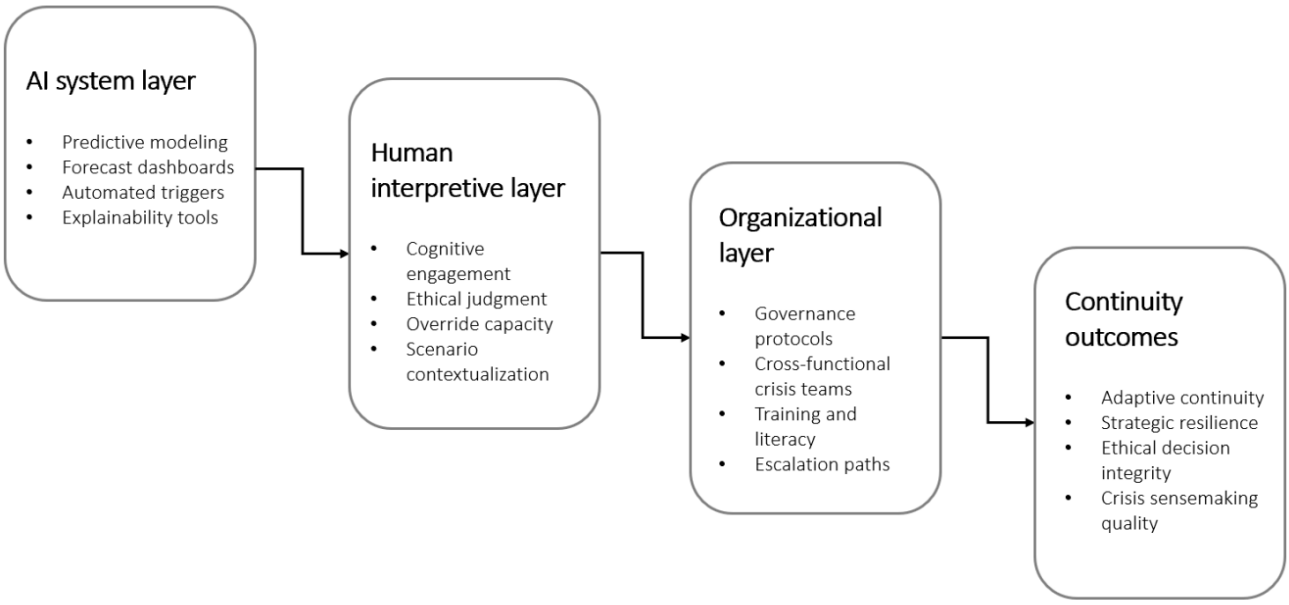
5.4. Conceptual Framework: Critical Interpretability and Resilient Continuity

From these findings, we derived a mid-range theoretical framework (Figure 1) linking AI system features, organizational enablers, and human cognitive practices to continuity outcomes. At its core lies the construct of critical interpretability, operationalized as a triadic interaction between:

1. AI affordances: including explainability, data provenance, and scenario simulation;
2. Organizational design: including culture, training, escalation structures, and interdepartmental collaboration;
3. Human interpretive work: including skepticism, contextual reasoning, ethical sensitivity, and override behaviors.

Where this triad is active and coherent, organizations exhibit adaptive continuity: faster response, better strategic alignment, and reduced error propagation. Where any element is weak or misaligned, fragile automation prevails, leading to delayed action, inappropriate responses, or ethical blind spots.

Figure 1: ***Critical Interpretability Framework for AI-Supported Business Continuity***



Source: ***our processing***

In particular, Figure 1 summarizes our empirical findings in a multi-layered framework of "*Critical Interpretability"* for AI-supported business continuity. It visualizes the dynamic interplay between AI system capabilities, human cognitive practices, and organizational conditions, and how these elements jointly shape continuity outcomes in crisis situations. The model begins with the AI System Layer, where data-driven tools, such as predictive analytics, automated triggers, and forecasting dashboards, provide rapid situational inputs. These tools, however, do not determine action by themselves. They must be interpreted, questioned, and situated by human actors operating within the Human Interpretive Layer. Here, decision-makers engage with AI outputs through ethical judgment, contextualization, and override decisions, often under severe time and moral pressure. Their ability to do so depends heavily on the Organizational Layer, which includes structures like cross-functional crisis teams, training protocols, and escalation pathways. These elements mediate how technology and cognition interact. When all three layers align, firms experience adaptive continuity, the capacity to respond quickly, ethically, and effectively to disruptive events. Conversely, disalignment leads to what we term fragile automation: continuity decisions that are technically rapid but contextually flawed or ethically blind. Thus, this framework is not merely a diagnostic tool, as it offers a new conceptual lens, *critical interpretability,* not as a system feature, but as a socio-cognitive capability, essential for resilience in increasingly AI-mediated environments.

6. Discussion

This study set out to examine how human interpretive judgment influences the effectiveness of AI-supported continuity planning in times of crisis. The analysis demonstrates that resilience does not stem from technological capability in isolation, but from the ongoing interaction between human cognition, ethical reasoning, and machine intelligence. What distinguishes adaptive organizations is not the sophistication of their algorithms, but their ability to interpret, question, and recontextualize what those algorithms produce. Through this lens, critical interpretability emerges as a foundational capability for

the age of intelligent uncertainty. It reframes explainability not as a property of systems but as a process of cognition—distributed, social, and ethically charged. Where traditional approaches to AI emphasize transparency and accountability mechanisms within the technology itself, our findings highlight the interpretive labor performed by humans who must render algorithmic logic intelligible within fluid and high-stakes contexts. In doing so, the study aligns with and extends the cognitive turn in crisis management, suggesting that sensemaking in AI-mediated environments must now account for hybrid cognitive architectures in which algorithms participate, but humans remain epistemic anchors. The discussion also reveals a paradox central to contemporary management: as organizations increasingly depend on AI for speed and analytical precision, they risk diminishing the very interpretive reflexivity that underpins resilience. Under conditions of time pressure and ambiguity, overreliance on automated insight can lead to what might be termed "*cognitive outsourcing*," where decision-makers defer to machine authority at the expense of contextual understanding. Conversely, organizations that sustain interpretive engagement—by cultivating ethical awareness, cognitive diversity, and cross-functional dialogue—transform AI from a decision instrument into a sensemaking partner. This relational view challenges the dominant discourse of automation as optimization. Rather than seeing AI as a substitute for human reasoning, it positions intelligence as a joint activity shaped by interaction, dialogue, and moral evaluation. In crisis situations, resilience thus depends not only on predictive accuracy or processing speed, but on the depth and quality of human interpretation—how actors discern relevance, surface blind spots, and reintegrate algorithmic outputs into evolving situational frames. In conceptual terms, critical interpretability extends beyond the technical vocabulary of explainable AI. It foregrounds the human capacities of reflexivity, skepticism, and moral judgment as constitutive elements of organizational intelligence. In doing so, it bridges the domains of cognitive management, AI ethics, and socio-technical resilience, offering a vocabulary for understanding how meaning is negotiated between humans and machines under pressure. Ultimately, the findings suggest that the future of AI-supported continuity will hinge on the ability of organizations to institutionalize this interpretive capability—to sustain not just data fluency but ethical fluency, not just explainability but comprehension. In an environment defined by uncertainty, resilience is thus less a matter of algorithmic power than of interpretive courage: the willingness to think critically with technology rather than through it.

6.1. Contributions

The first major contribution of this study lies in the conceptualization of *critical interpretability* as a hybrid cognitive–organizational capability. Rather than treating interpretability as a technical attribute of algorithms, this research positions it as a situated human practice—a dynamic process that unfolds through reflection, ethical reasoning, and contextual adaptation. Interpretability, in this sense, varies across crises and organizational environments, shaped by culture, structure, and the degree of trust between human and machine actors. The second contribution concerns the understanding of business continuity as a relational and interpretive process rather than a procedural or digital one. Continuity is not achieved through pre-defined checklists or automated dashboards, but through judgment exercised under uncertainty—where decision-makers must make sense of incomplete information, weigh moral trade-offs, and narratively reframe algorithmic signals into coherent courses of action. This perspective repositions continuity as a practice of meaning-making, in which human cognition and AI capabilities converge to sustain organizational responsiveness. Third, the study offers an integrated, multi-level framework linking individual cognitive engagement with organizational enablers and systemic outcomes. By connecting micro-level interpretive behaviors to meso-level governance structures and macro-level resilience performance, the framework explains why organizations facing similar technological conditions display very different capacities for adaptation and recovery. Collectively, these contributions address the research questions that guided the study. The first question—how human engagement with AI systems affects continuity—finds its answer in the observation of interpretive actions such as contextual reframing, override behavior, and ethical intervention. The second question—what cognitive and organizational factors enable effective interpretation—is addressed through the identification of enablers including AI literacy, cross-functional coordination, and structured escalation pathways. The third question—what

systemic conditions support human–AI integration—emerges through the analysis of governance practices, trust architectures, and socio-technical coherence. In fulfilling its objectives, the study accomplishes four interrelated outcomes:

1. It examines interpretive practices in depth, revealing how judgment is exercised in AI-mediated crises.

2. It identifies the conditions that enhance or inhibit human–AI collaboration.

3. It develops and refines the construct of *critical interpretability* as a distinct managerial capability.

4. It proposes a structured, conceptually grounded framework that clarifies the cognitive, ethical, and organizational foundations of resilient continuity.

6.2. Managerial and Policy Implications

For managers, the findings of this study make clear that AI cannot be approached as a plug-and-play solution for continuity planning or crisis response. The effectiveness of AI depends not on its algorithmic precision, but on the human and organizational infrastructures that sustain interpretive engagement. Building resilience therefore requires deliberate investment in training programs that strengthen cognitive preparedness, ethical awareness, and interpretive confidence. Managers must learn not only how to read data, but how to question, reframe, and contextualize algorithmic outputs within the unfolding realities of a crisis. Organizations should design governance systems that recognize override authority as a strategic safeguard rather than a failure of automation. This means empowering individuals and teams to pause or adjust AI recommendations when situational evidence, stakeholder concerns, or ethical ambiguities emerge. Such empowerment transforms human judgment from a reactive fallback into a proactive element of continuity management. Similarly, cross-functional crisis teams that combine technical, operational, and ethical expertise can translate predictive insights into coordinated action, ensuring that AI serves as an integrative rather than fragmenting force. A further managerial risk identified in this research is automation complacency—the tendency to treat AI systems as objective, error-free arbiters of truth. When decision-makers place excessive trust in automated logic, they may overlook context, nuance, or weak signals that fall outside predefined parameters. This cognitive overreliance can be especially dangerous under crisis conditions, when data are incomplete and time is compressed. Training managers to engage critically with AI, to challenge and reinterpret its recommendations, becomes therefore not a liability but a core resilience asset. From a policy perspective, these findings suggest that AI governance must evolve beyond narrow technical standards of accuracy and performance. Effective governance in high-stakes sectors such as energy, health, and finance requires a parallel focus on organizational interpretability readiness—the institutional capacity to embed ethical reasoning, transparency, and human oversight into AI deployment. This includes the establishment of internal AI ethics boards, mandatory audit and escalation procedures, and participatory design processes that involve end-users in system development. Regulatory frameworks should encourage not only compliance, but capability building: policies that foster digital literacy, interdisciplinary collaboration, and shared accountability between human and algorithmic agents. In doing so, governance moves from a rule-based paradigm to a learning-oriented paradigm, where firms and regulators co-evolve interpretive competence in tandem with technological advancement. Ultimately, both managerial and policy actors share responsibility for shaping an ecosystem in which AI enhances rather than erodes human judgment. The challenge is not to automate continuity, but to cultivate the interpretive infrastructures—cognitive, ethical, and institutional—that make continuity genuinely intelligent.

6.3. Limitations and Future Research

While this study provides a rich empirical and conceptual foundation, it is necessarily bounded by its context. The focus on Italian organizations and the crisis window of 2020–2024 offers depth and specificity but limits generalizability. Future research should extend the analysis to multinational and cross-cultural settings, exploring how interpretive practices vary across governance systems, regulatory environments, and technological infrastructures. Comparative studies could reveal whether critical interpretability manifests differently in centralized versus networked organizations, or in public versus private sectors. A second avenue lies in longitudinal research tracing how interpretive capabilities evolve over time. As organizations integrate AI more deeply into strategic and operational decision-making, interpretability may shift from being an emergent practice to an institutionalized competence—embedded in governance routines, training systems, and digital ethics protocols. Measuring this evolution could clarify the relationship between interpretability maturity and resilience outcomes. Quantitative studies could complement qualitative insights by developing scales and indicators that assess interpretive readiness, cognitive diversity, and ethical reflexivity within AI-enabled organizations. This would help establish empirical links between interpretability and performance metrics such as recovery speed, error mitigation, and adaptive capacity. Equally important, future work should examine how emerging AI paradigms—including generative systems, large-scale autonomous agents, and self-learning organizational platforms—reshape the interpretability landscape. As AI systems become capable of not only analyzing but generating strategic options, the boundary between decision support and decision authority will blur. This raises pressing questions: How can human judgment be preserved or even amplified in the face of increasing machine autonomy? What new governance models will ensure that foresight and ethics remain central to decision-making when AI participates in strategy formulation itself? Finally, we call for an integrative research agenda linking cognitive science, organizational theory, and AI ethics to study interpretability as a form of collective intelligence. Such an agenda would move beyond individual cognition to explore how teams, institutions, and technologies co-produce meaning under pressure. This shift—from human versus machine to human-with-machine—represents the next frontier for resilience scholarship. At a moment when crises are becoming more complex and AI systems increasingly pervasive, the challenge is not simply to automate continuity but to sustain human intelligence within intelligent systems: the future of resilience will depend on designing socio-technical architectures that enable humans to interpret, challenge, and ethically steer AI, transforming automation into augmented judgment.

References

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS Quarterly*, Association for Information Systems.

Boin, A., & van Eeten, M. J. G. (2013). The resilient organization. *Public Management Review*, Taylor & Francis.

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, Wiley.

Chatterjee, S., Rana, N. P., Tamilmani, K., & Sharma, A. (2021). The next generation of smart business: A bibliometric and visualization analysis. *Technological Forecasting and Social Change*, Elsevier.

Dignum, V. (2018). *Responsible Artificial Intelligence: Designing AI for human values*. Springer, Cham.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. Cornell University.

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2021). Re-examining the implementation of artificial intelligence (AI): AI readiness, assimilation, and technological capability in organizations. *International Journal of Information Management*, Elsevier.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, Elsevier.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, Academy of Management.

Grundner, C., & Neuhofer, B. (2021). Human-AI interaction in high-stress domains: The role of explainability and interactivity. *Journal of Decision Systems*, Taylor & Francis.

Hodgkinson, G. P., & Healey, M. P. (2011). Psychological foundations of dynamic capabilities: Reflexion and reflection in strategic management. *Strategic Management Journal*, Wiley.

Ivanov, D., & Dolgui, A. (2020). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control*, Taylor & Francis.

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. *University of Pennsylvania Law Review*, University of Pennsylvania Press.

Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. *Academy of Management Annals*, Academy of Management.

Mikalef, P., Krogstie, J., & Pappas, I. O. (2022). Ethical AI design in high-uncertainty organizational contexts: A multi-level perspective. *Information & Management*, Elsevier.

Papagiannidis, S., See-To, E. W. K., & Bourlakis, M. (2023). Human-AI collaboration risks: Bias, deskilling, and blind spots in AI-driven operations. *Journal of Business Research*, Elsevier.

Paschen, J., Wilson, M., & Ferreira, J. (2020). Artificial intelligence (AI) and its implications for market intelligence in crisis management. *Journal of Business Research*, Elsevier.

Power, M. J., Makel, M. C., & Kazda, G. (2021). Decision-making under pressure: Human-AI teaming in crisis response. *International Journal of Crisis Communication*, CrisisComm Publications.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, Springer.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, Academy of Management.

Tarafdar, M., Beath, C., & Ross, J. (2019). Understanding digital resilience: Artificial intelligence in sensing and responding to crises. *Information Systems Frontiers*, Springer.

Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly*, SAGE Publications.

Weick, K. E., & Sutcliffe, K. M. (2015). *Managing the Unexpected: Sustained Performance in a Complex World* (3rd ed.). Wiley, Hoboken.

Williams, T. A., Gruber, D. A., Sutcliffe, K. M., Shepherd, D. A., & Zhao, E. Y. (2017). Organizational response to adversity: Fusing crisis management and resilience research streams. *Academy of Management Annals*, Academy of Management.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage Publications, Thousand Oaks, CA.