© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718609

AI-Powered Disaster Recovery Planning and Failover Optimization for Mission-Critical Enterprise Systems

TAHSEEN ZAFAR

Abstract- ***Mission-critical enterprise systems now operate through hybrid combinations of private data centres, public cloud zones, edge services, identity platforms, databases, application programming interfaces, and third-party digital supply chains. Conventional disaster recovery planning, while still necessary, is increasingly insufficient because static runbooks cannot anticipate fast-moving failures, cascading dependencies, ransomware disruptions, configuration drift, and volatile workload peaks. This review examines how artificial intelligence can strengthen disaster recovery planning and failover optimization through predictive failure analysis, anomaly detection, automated root cause reasoning, workload forecasting, policy-aware orchestration, and human-supervised remediation. A structured narrative review method was applied to recent research and standards published between 2020 and 2025, with emphasis on AIOps, cloud reliability, critical infrastructure protection, generative artificial intelligence, autoscaling, microservice observability, cyber resilience, and governance. The paper proposes an integrated AI-driven disaster recovery lifecycle that links telemetry ingestion, risk scoring, scenario modelling, failover decisioning, validation, and post-incident learning. The review indicates that the strongest value of AI lies not in replacing continuity professionals but in compressing detection-to-decision time, reducing alert noise, identifying probable blast radius, and recommending recovery actions aligned with recovery time objectives, recovery point objectives, security controls, and service-level commitments. However, the literature also reveals persistent limitations, including explainability gaps, poor data quality, adversarial model risk, over-automation, weak integration with legacy platforms, and uncertain accountability during autonomous failover. The study concludes that AI-powered disaster recovery should be designed as a governed socio-technical capability, combining machine intelligence, resilient architecture, audited automation, and expert approval for high-impact actions.***

Keywords: Artificial Intelligence, Disaster Recovery Failover Optimization AIOps, Mission-Critical Systems, Cyber Resilience, Business Continuity.

I. INTRODUCTION

Enterprise resilience has become a strategic concern because digital services now mediate finance, healthcare, logistics, energy, government, manufacturing, education, and customer interaction. Mission-critical systems are no longer isolated applications running inside a single data centre; they are distributed portfolios of microservices, virtual machines, containers, databases, message brokers, identity services, cloud regions, backup repositories, and external software dependencies. A disruption in one component can propagate across authentication, payment, reporting, analytics, and customer channels. Recent work on critical infrastructure protection shows that cyberattacks, cloud outages, industrial control threats, and operational failures increasingly require integrated incident response and business continuity planning, rather than separate security and recovery playbooks [1].

Traditional disaster recovery planning is founded on impact analysis, risk assessment, backup design, alternate site provisioning, recovery time objectives, recovery point objectives, runbooks, and testing. These foundations remain essential, yet they struggle in environments where telemetry volume exceeds human attention and dependencies change several times per day. Cloud-native and microservice systems generate logs, metrics, traces, events, change records, user journeys, configuration states, and security alerts at high velocity. Surveys of cloud and microservice reliability emphasize that anomaly detection and root cause analysis have become central to rapid recovery because human operators cannot manually interpret every symptom under pressure [5]. AIOps research extends this argument by showing that artificial intelligence can structure operational knowledge, correlate signals, and support faster diagnosis during incidents [7,16].

The reviewed literature suggests a shift from reactive recovery to anticipatory resilience. Instead of waiting for service collapse, AI-enabled platforms can learn normal behaviour, forecast resource exhaustion, detect abnormal patterns, estimate component failure probability, and initiate pre-approved protection actions. Deep reinforcement learning and predictive autoscaling studies show that dynamic policies can allocate resources more effectively than static thresholds in uncertain cloud environments [9-13]. Similarly, generative AI and large language models can help summarize incident timelines, classify alerts, produce recovery hypotheses, and assist operators with domain-specific reasoning, although they require strong guardrails because hallucinated recommendations are unacceptable in mission-critical operations [1,2,6].

This paper reviews AI-powered disaster recovery planning and failover optimization for mission-critical enterprise systems. Its aim is to develop a coherent synthesis of recent research and practice, explain the mechanisms through which AI can improve recovery, identify risks that could undermine safe adoption, and propose a review-based framework suitable for enterprise planning. The objectives are: first, to map AI functions across the disaster recovery lifecycle; second, to examine failover optimization methods for hybrid and cloud-native systems; third, to identify governance requirements for trustworthy automation; fourth, to compare relevant technical approaches from the 2020 to 2025 literature; and fifth, to highlight open research questions for resilient, auditable, and explainable AI-assisted recovery.

II. REVIEW METHODOLOGY

This article uses a structured narrative review methodology appropriate for an emerging interdisciplinary topic that spans disaster recovery, AIOps, cloud reliability, cybersecurity, and critical infrastructure protection. The review focused on studies, technical standards, and authoritative guidance published from 2020 to 2025. Search concepts included artificial intelligence for IT operations, cloud disaster recovery, autonomous failover, predictive autoscaling, root cause analysis, incident response, cyber resilience, critical infrastructure, large language models, and business continuity. The inclusion criteria prioritised sources that addressed at least one of four areas: failure prediction and anomaly detection; automated or semi-automated remediation; cloud or microservice resource optimization; and governance of AI-enabled security or continuity controls. Sources were excluded when they were unrelated to enterprise recovery, focused solely on consumer backup, or did not provide transferable technical or governance insight.

The review adopted thematic synthesis rather than statistical aggregation because the field is heterogeneous. Some studies evaluate algorithms on telemetry datasets, others propose architectures, and others provide standards for security governance. Evidence was grouped into six themes: observability and early warning; root cause analysis and knowledge extraction; workload forecasting and resource allocation; failover orchestration; security, privacy, and regulatory assurance; and human oversight. The attached critical infrastructure review informed the framing of reliability, response speed, trust, privacy, and resilience, especially its emphasis on linking incident response with business continuity and evaluating AI in high-consequence domains [1]. Cloud security literature informed the threat model, including data breaches, insecure interfaces, malicious insiders, identity compromise, and multitenancy risks that remain relevant during disaster recovery operations [4].

The methodology has three strengths. It captures recent technical development without forcing incompatible studies into a single quantitative model. It integrates academic research with operational standards, which is necessary because disaster recovery is both an engineering and a governance discipline. It also allows the paper to identify design principles that can guide organisations even when they use different cloud providers, databases, and enterprise platforms. Its main limitation is that many AI-driven recovery tools are proprietary, meaning that public evidence may under-represent industrial deployments. Therefore, the proposed framework is positioned as a review-based synthesis, not as a validated product benchmark.

2.1 Review questions and analytical lens

The review was guided by five questions. How can AI identify failures earlier than conventional threshold monitoring? Which forms of intelligence are most useful for failover decision-making under uncertainty? How should enterprises balance automation speed with accountability? What technical and organisational barriers prevent AI-enabled recovery from achieving dependable results? Which research directions could make autonomous or semi-autonomous recovery safer for regulated and high-availability environments? These questions were selected because disaster recovery involves more than backup restoration; it requires a coordinated response across technology, people, process, contracts, and risk appetite.

The analytical lens used in this study combines reliability engineering, cyber resilience, and operational learning. Reliability engineering focuses on avoiding and recovering from component failure. Cyber resilience focuses on maintaining essential functions despite malicious compromise. Operational learning focuses on how incidents, rehearsals, and telemetry improve future readiness. This combined lens is necessary because an AI model that predicts a server fault but ignores data integrity, access control, or business priority cannot be considered a complete recovery tool. Likewise, a secure recovery design that lacks observability may fail to activate quickly enough. The synthesis therefore evaluates AI methods according to four criteria: early warning value, actionability, controllability, and auditability.

2.2 Scope of mission-critical enterprise systems

For this review, mission-critical enterprise systems are defined as digital services whose prolonged unavailability would cause severe financial loss, legal exposure, safety risk, reputational damage, or interruption of essential organisational functions. Examples include payment switching, core banking ledgers, hospital clinical platforms, national service portals, enterprise resource planning, industrial control gateways, airline reservation systems, logistics platforms, and customer identity infrastructure. These systems share three characteristics. They are dependency-rich, because many upstream and downstream services must operate together. They are data-sensitive, because availability must be balanced against integrity and confidentiality. They are time-constrained, because recovery windows are measured in minutes or hours rather than days.

The scope includes on-premises, cloud, edge, and hybrid deployments because modern continuity programmes rarely operate in a single environment. A database may remain on-premises for regulatory or latency reasons while applications run in containers across multiple cloud zones. Backup storage may be immutable and remote, while security monitoring is delivered as a managed service. Identity may rely on external federation, and communication channels may rely on third-party networks. AI-powered disaster recovery must therefore reason across platforms, not only within one cloud console. This is why dependency mapping, policy alignment, and human communication remain as important as algorithms.

2.3 Quality assessment of evidence

The reviewed sources were assessed for relevance, recency, methodological transparency, and transferability. Algorithmic studies were valued when they explained datasets, evaluation metrics, assumptions, and limitations. Standards and guidance were valued when they provided actionable controls for governance, security, continuity, or AI risk. Review papers were valued when they clarified taxonomies or compared techniques across multiple studies. The strongest evidence came from sources that linked technical performance with operational context, such as microservice anomaly detection, cloud incident root cause analysis, and critical infrastructure protection [1,5,8,15].

Some caution was necessary. Reported accuracy improvements in laboratory settings may not translate directly into production recovery because real incidents include incomplete telemetry, organisational stress, vendor dependencies, legal concerns, and unusual failure combinations. Similarly, an AI model that performs well on historical incidents may fail when a new architecture, a zero-day exploit, or a cloud provider outage changes the incident pattern. Therefore, this review treats AI as a decision support and optimization

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718609

capability that must be validated continuously rather than as a one-time technology purchase.

2.4 Synthesis framework

The synthesis framework developed from the literature consists of five layers. The data layer collects metrics, logs, traces, events, configuration data, backup status, security alerts, and business context. The intelligence layer applies anomaly detection, forecasting, dependency analysis, natural language processing, and risk scoring. The decision layer ranks recovery options against objectives and constraints. The orchestration layer executes or recommends actions such as scaling, traffic shifting, replica promotion, workload migration, backup restoration, and graceful degradation. The governance layer supervises identity, approval, audit, explainability, compliance, testing, and continuous learning.

This layered structure prevents a common implementation mistake: treating AI as a stand-alone monitoring feature rather than a recovery management capability. A model may detect an anomaly, but disaster recovery value is created only when the detection is connected to dependency understanding, feasible action, validated restoration, and accountable review. The framework also supports staged maturity. Organisations can begin with passive recommendations, then progress to low-risk automation, and finally permit controlled failover actions after evidence demonstrates reliability.

2.5 Evaluation logic for review conclusions

The conclusions of this review were derived through cross-theme comparison. Each AI capability was examined against the disaster recovery outcome it claimed to improve. Failure prediction was linked to earlier activation; root cause analysis was linked to faster diagnosis; resource optimization was linked to capacity readiness; generative summarization was linked to operator understanding; and automated orchestration was linked to reduced execution delay. Capabilities that did not clearly improve a recovery objective were treated as supporting tools rather than core mechanisms.

The evaluation also considered negative evidence. AI can create new fragility when models are trained on biased incident histories, when telemetry pipelines fail during the same event they are supposed to diagnose, or when automation expands a configuration error across regions. In addition, some recovery actions involve irreversible business consequences, such as accepting data loss, promoting a secondary database, or isolating a network segment during suspected compromise. These decisions require policy, ownership, and escalation rules. For this reason, the review gives equal attention to governance and optimization. A technically elegant failover model is not acceptable if it cannot explain its recommendation, preserve forensic evidence, respect legal obligations, and support rollback.

The final synthesis was therefore judged against six practical questions: Does the capability reduce uncertainty before failure? Does it support an action within the required recovery window? Does it protect data integrity and security? Can its recommendation be explained to engineers and auditors? Can it be tested without harming production systems? Does it improve after incidents and exercises? These questions provide a bridge between academic AI performance metrics and enterprise continuity requirements. They also support a cautious but progressive adoption path in which organisations use AI first to observe and advise, then to automate routine stabilization, and only later to initiate high-impact failover under strict approval.

Figure 1. Integrated lifecycle for AI-powered disaster recovery planning and governed failover optimization.

III. CONCEPTUAL BASIS OF AI-POWERED DISASTER RECOVERY

AI-powered disaster recovery can be defined as the use of machine learning, knowledge representation, generative reasoning, and automated decision support to improve the preparation, activation, execution, and learning phases of recovery. In mission-critical environments, the purpose is not merely to restore servers; it is to preserve essential business outcomes under degraded conditions. This requires awareness of service dependencies, data consistency, cyber risk, infrastructure capacity, legal obligations, customer impact, and operational priority. AI contributes by turning telemetry and historical incident records into actionable signals before, during, and after disruption.

The first contribution is predictive awareness. Time-series forecasting, anomaly detection, and behaviour modelling can estimate whether a workload, storage cluster, network path, or database replication channel is moving toward failure. Microservice studies demonstrate that logs, metrics, and distributed traces provide complementary evidence; when combined, they can detect performance anomalies and support localization [5,14,15]. This matters for disaster recovery because failover is most effective when triggered before queues saturate, backups fall behind, or replicas diverge.

The second contribution is decision optimization. Failover decisions are constrained by recovery time objectives, recovery point objectives, cost, capacity, geographic separation, compliance requirements, and security posture. Reinforcement learning and predictive autoscaling research indicate that AI agents can learn policies for resource allocation under dynamic demand, where static threshold rules often underperform [10,11]. In disaster recovery, the same principle can be adapted to select whether the system should scale locally, shift traffic to a standby zone, promote a replica, activate an alternate cloud region, degrade non-essential functions, or invoke manual approval.

The third contribution is knowledge acceleration. Incident records often contain valuable explanations, but they are written in unstructured language. Studies on mining root cause knowledge from cloud service investigations and applying large language models to cloud incidents show that natural language processing can extract causation patterns, retrieve similar incidents, and generate diagnostic narratives for engineers [6,7]. This capability is useful during failover because teams need rapid access to what previously worked, what failed, and which dependencies were affected.

The fourth contribution is continuous learning. Disaster recovery plans often decay because applications change, dependencies drift, and recovery assumptions are not retested. AI can compare runbooks against observed architecture, detect missing owners, flag untested dependencies, and recommend scenario updates. Nevertheless, AI outputs must be reviewed, version-controlled, and tested. Standards for cybersecurity and AI governance stress risk management, accountability, documentation, and monitoring across the lifecycle [20-25].

IV. AI FUNCTIONS ACROSS THE DISASTER RECOVERY LIFECYCLE

The disaster recovery lifecycle can be divided into preparation, detection, decision, execution, validation, and learning. AI can support each phase when suitable controls are implemented. During preparation, machine learning can analyse historical incidents, asset inventories, dependency maps, change tickets, vulnerability data, capacity histories, and business impact ratings. The goal is to identify which services require hot standby, warm standby, asynchronous replication, immutable backup, or manual restoration. Cloud security research shows that confidentiality, access control, authentication, auditing, and privacy must remain part of this preparation, because recovery platforms can become attractive targets for attackers [4].

During detection, AI models monitor streams of telemetry and identify deviations from normal behaviour. Distributed tracing and log-based techniques are especially valuable because they capture service interactions that are invisible in isolated metrics [15,17-19]. However, detection must be tuned to business criticality. A minor latency spike in an internal reporting service should not trigger a costly regional failover, whereas a replication lag in a payment database may require immediate escalation.

Thus, AI detection must feed a risk model that includes technical severity, service tier, customer impact, and data loss exposure.

During decisioning, failover optimization evaluates candidate actions. A useful AI model should consider present health, predicted future health, recovery objectives, network constraints, cost, dependency readiness, and confidence. For example, a model may recommend traffic shifting if the primary zone shows correlated network degradation, the standby zone is healthy, data replication lag is inside the recovery point objective, and identity services are reachable. Conversely, it may recommend graceful degradation when failover would move load to an under-provisioned region. Markov models and reliability analysis provide useful ways to reason about operational and failed states, while AI can estimate transition probabilities from live data [1].

During execution, automation should follow pre-approved guardrails. Low-risk actions, such as scaling read replicas or increasing queue consumers, may be executed automatically. Medium-risk actions, such as draining traffic from one availability zone, may require on-call confirmation. High-risk actions, such as database promotion, cross-region write enablement, or ransomware recovery from immutable backup, should require explicit multi-party approval. This tiered approach avoids the danger of over-automation while still reducing response time.

During validation, AI can compare expected and observed recovery states. It can check service availability, data freshness, transaction success, security control continuity, access policy consistency, and customer experience. After the incident, learning systems can update runbooks, improve failure signatures, revise capacity assumptions, and identify architecture weaknesses. This closes the loop between disaster recovery planning and operational reliability.

Table 1. Thematic evidence matrix for AI-powered disaster recovery research from 2020 to 2025.

Theme	Recent evidence base	Implication for recovery planning	Gap requiring attention
Early warning	Microservice anomaly detection, tracing, logs, and distributed AI [5,14,15,17-19]	Detect degradation before complete outage and link alerts to service context.	Models fail when telemetry is incomplete or instrumented inconsistently.
Diagnosis	Cloud RCA, incident knowledge graphs, and LLM incident analysis [6-8]	Shortens investigation by retrieving similar cases and summarising evidence.	Generated explanations require validation against authoritative telemetry.
Optimization	RL autoscaling and cloud-edge resource allocation [9-13]	Improves pre-failover capacity allocation under changing workload pressure.	Training policies safely for rare disasters remains difficult.
Governance	AI risk, cybersecurity, and security management standards [20-25]	Creates auditability, accountability, and lifecycle controls for AI-assisted actions.	Standards must be translated into executable recovery guardrails.
Operational practice	Cloud reliability and recovery guidance [26-30]	Supports testing, resilience architecture, and cross-functional continuity ownership.	Provider-specific guidance can be difficult to harmonise across hybrid estates.

V. FAILOVER OPTIMIZATION IN HYBRID ENTERPRISE SYSTEMS

Failover optimization is the process of selecting the safest and fastest transition from a degraded primary environment to an alternate operating mode. In hybrid enterprise systems, the challenge is intensified by mixed infrastructure: on-premises databases, cloud replicas, software-as-a-service dependencies, identity federations, virtual private networks, edge gateways, and legacy applications. A failover action

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718609

that restores one application may break another if shared dependencies are not understood.

AI can improve failover optimization by constructing a dynamic dependency graph. Nodes represent services, databases, infrastructure components, network links, data stores, security controls, and business processes. Edges represent calls, data replication, authentication, message exchange, or operational ownership. Telemetry updates the graph in near real time. Graph-based reasoning then estimates blast radius, identifies critical paths, and highlights single points of failure. Recent microservice and cloud reliability studies support the use of topology, logs, and key performance indicators for root cause inference, particularly when systems are too complex for manual reasoning [8].

Optimization should also include objectives that sometimes conflict. The fastest failover may not be the safest if data replication is incomplete. The lowest-cost recovery may not meet service-level commitments. The most automated response may violate change-control requirements. Therefore, AI-powered failover must be multi-objective. A scoring function can combine expected recovery time, expected data loss, confidence, security posture, compliance constraints, cost, and customer impact. Reinforcement learning can help tune policies under simulated workloads, but live enterprise use should begin with recommendation mode and progress gradually as confidence increases [10-13].

A practical failover score can be expressed as a weighted decision index: F = w1RTO + w2RPO + w3C + w4S + w5B - w6K, where RTO is predicted recovery time, RPO is predicted data loss, C is capacity risk, S is security exposure, B is business impact, and K is confidence in the recommendation. The weights should be approved by business and technical owners. This model is not intended to replace engineering judgement; it provides transparent reasoning that can be audited after the event.

Figure 2. Risk-scored failover decision flow with controlled recovery action tiers.

VI. SECURITY AND GOVERNANCE REQUIREMENTS

AI-powered disaster recovery introduces security risks as well as benefits. Recovery systems hold credentials, backup keys, replication channels, privileged automation scripts, network configurations, and recovery images. If compromised, they can become a mechanism for destructive attack. Cloud security reviews identify risks such as account hijacking, insecure interfaces, unauthorized access, key exposure, weak auditing, and abuse of cloud services [4]. During a disaster, these risks increase because teams may bypass controls to restore service quickly.

Governance must therefore be embedded into every AI-enabled recovery workflow. Identity and access management should enforce least privilege, separation of duties, just-in-time access, and tamper-resistant audit logs. Backup repositories should support immutability, encryption, malware scanning, and independent credentials. AI models should be protected against data poisoning, prompt injection, insecure tool execution, and unauthorized access to sensitive telemetry. Reviews of generative AI in cybersecurity warn that large language models can help defenders but also introduce vulnerabilities, including unsafe outputs and adversarial manipulation [2,3,23].

Explainability is essential. Operators should know why a failover recommendation was produced, which evidence was used, how confident the model is, and what risks remain. For regulated industries, auditability is not optional. Frameworks such as the

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718609

AI Risk Management Framework, the Cybersecurity Framework, and information security management standards emphasize governance, measurement, documentation, and continuous improvement [20-25]. Applied to disaster recovery, this means that model changes, automation permissions, runbook updates, and failover decisions should be logged and reviewed.

Human oversight should be risk-based. It is inefficient to require manual approval for every self-healing action, but it is unsafe to permit unsupervised high-impact failover. A balanced model defines action tiers, approval thresholds, rollback criteria, and emergency override rules. The most mature organisations will treat AI as an operational co-pilot: fast at correlation, memory, and simulation, but constrained by policy and accountable human judgement.

Table 2. Evaluation metrics and governance controls for AI-enabled failover optimization.

Metric/control	Purpose	AI-enabled use	Governance safeguard
Predicted RTO	Estimated time to restore a service tier.	Ranks candidate actions by expected recovery speed.	Compare predicted and observed RTO after every exercise.
Predicted RPO	Estimated data loss or replica lag.	Blocks failover when data freshness is outside policy.	Require data owner approval for exception handling.
Blast radius score	Likely propagation across dependent services.	Prioritises isolation, traffic shifting, and dependency checks.	Maintain a validated dependency graph and audit changes.
Confidence level	Model certainty based on evidence quality.	Determines whether advice, approval, or automation is permitted.	Set minimum confidence thresholds by action tier.
Security posture	Trustworthiness of target recovery environment.	Checks identity, vulnerability, backup integrity, and malware signals.	Separate recovery credentials and immutable logs.
Rollback readiness	Ability to reverse or contain a recovery action.	Selects actions with tested rollback paths under uncertainty.	Exercise rollback plans and document exceptions.

VII. DISCUSSION

The reviewed literature converges on one central insight: AI improves disaster recovery when it is connected to reliable observability, architectural knowledge, and governed automation. It is less effective when deployed as a generic prediction layer over poor telemetry or outdated runbooks. The dependency between data quality and recovery quality is particularly important. If logs lack service identifiers, traces are incomplete, asset inventories are stale, or ownership data is missing, AI may produce plausible but unreliable recommendations.

For mission-critical enterprise systems, the most promising near-term use cases are anomaly detection, alert correlation, incident summarization, similarity search across past incidents, capacity forecasting, recovery readiness scoring, and decision support for failover. These use cases reduce cognitive overload without requiring full autonomy. More advanced use cases include automated traffic shifting, pre-emptive replica scaling, reinforcement-learning-based resource allocation, and generative runbook updating. These should be introduced through simulation, controlled pilots, and staged approvals.

The review also highlights the importance of cyber resilience. Disaster recovery is no longer concerned only with hardware failure or natural disasters. Ransomware, destructive malware, cloud credential compromise, software supply chain attacks, insider abuse, and data corruption can all require recovery.

AI systems must therefore distinguish between ordinary failure and adversarial disruption. A failover triggered by ransomware could replicate corrupted data if not governed by integrity checks and clean recovery points. Conversely, delayed failover during a genuine service outage can breach contractual commitments. The right response depends on accurate classification and risk-aware decisioning.

Another issue is organisational readiness. AI-powered recovery requires cooperation between continuity managers, site reliability engineers, cybersecurity teams, data owners, application teams, legal staff, and executive sponsors. Without shared ownership, the model may optimize technical metrics while neglecting business priorities. For example, a system might prioritize restoring a high-traffic application, while the business may require a lower-traffic regulatory reporting function to recover first. Therefore, business impact analysis should provide input to model design.

VIII. IMPLICATIONS FOR ENTERPRISE IMPLEMENTATION

Enterprises adopting AI-powered disaster recovery should begin by strengthening foundations. First, they should classify services by criticality, recovery objectives, data sensitivity, dependency complexity, and regulatory exposure. Second, they should unify observability so that logs, metrics, traces, synthetic tests, configuration changes, backup status, security events, and customer experience signals can be correlated. Third, they should build dependency maps that remain current through automated discovery and human validation.

Fourth, enterprises should define failover policies before incidents occur. Policies should specify which actions can be automated, which require approval, and which are prohibited without executive authorization. Fifth, they should test AI recommendations through tabletop exercises, chaos experiments, backup restoration drills, cyber recovery simulations, and controlled failover rehearsals. Sixth, they should measure outcomes using mean time to detect, mean time to decide, mean time to recover, restoration accuracy, data loss, false positive rate, false negative rate, operator workload, and customer impact.

Implementation should be incremental. A sensible maturity path starts with AI-assisted observability and incident summarization, then moves to recovery readiness scoring, then semi-automated low-risk remediation, and finally controlled failover automation. At each step, organisations should confirm that the model improves decisions, not just dashboards. The evidence should be measurable, repeatable, and stakeholder reviewed across critical operating contexts [26-30]. Together, these measures make AI recovery credible: proactive, evidence-led, transparent, restrained, rehearsed, and aligned with the real priorities of enterprise resilience under severe disruption and sustained digital trust globally.

IX. LIMITATIONS AND FUTURE RESEARCH

The current literature offers strong evidence for AI-assisted detection, diagnosis, and resource optimization, but less public evidence for fully autonomous disaster recovery in large regulated enterprises. Many real-world deployments remain confidential. There is also limited research on how generative AI should be validated for recovery recommendations that involve legal, financial, or safety consequences. Future studies should evaluate AI-driven failover in realistic hybrid environments with databases, identity services, network segmentation, backup repositories, and cyberattack scenarios.

Further research is needed on explainable failover scoring, adversarially robust recovery models, synthetic incident datasets, cross-cloud recovery simulation, and integration between business impact analysis and AI optimization. The field would also benefit from benchmarks that measure not only detection accuracy but also safe action selection, data consistency, rollback success, and operator trust.

X. CONCLUSION

AI-powered disaster recovery planning and failover optimization can materially improve the resilience of mission-critical enterprise systems when implemented with disciplined governance. The best

results are likely to come from systems that combine predictive analytics, microservice observability, root cause knowledge, workload forecasting, dependency graphs, and policy-controlled automation. AI can reduce detection time, support faster diagnosis, recommend safer failover options, and keep recovery plans aligned with changing architecture. Yet it cannot remove the need for tested backups, resilient design, clear ownership, security controls, and expert judgement. The most defensible model is human-supervised intelligence: machines correlate and recommend at speed, while accountable professionals define objectives, approve high-impact actions, and learn from each incident. In this form, AI becomes a practical accelerator of continuity, not an ungoverned substitute for resilience engineering.

REFERENCES

[1] Yigit Y, Ferrag MA, Ghanem MC, Sarker IH, Maglaras LA, Chrysoulas C, Moradpoor N, Tihanyi N, Janicke H (2025) Generative AI and LLMs for critical infrastructure protection: evaluation benchmarks, agentic AI, challenges, and opportunities. Sensors 25(6):1666. https://doi.org/10.3390/s25061666

[2] Ferrag MA, Alwahedi F, Battah A, Cherif B, Mechri A, Tihanyi N, Debbah M, Lestable T (2025) Generative AI in cybersecurity: a comprehensive review of LLM applications and vulnerabilities. Internet of Things and Cyber-Physical Systems 5:1-46. https://doi.org/10.1016/j.iotcps.2025.01.001

[3] Uddin M, Irshad MS, Kandhro IA, Alanazi F, Ahmed F, Maaz M, Hussain S, Ullah SS (2025) Generative AI revolution in cybersecurity: a comprehensive review of threat intelligence and operations. Artificial Intelligence Review 58:236. https://doi.org/10.1007/s10462-025-11219-5

[4] Dawood M, Tu S, Xiao C, Alasmary H, Waqas M, Rehman SU (2023) Cyberattacks and security of cloud computing: a complete guideline. Symmetry 15(11):1981. https://doi.org/10.3390/sym15111981

[5] Soldani J, Brogi A (2022) Anomaly detection and failure root cause analysis in (micro)service-based cloud applications: a survey. ACM Computing Surveys 55(3):1-39. https://doi.org/10.1145/3501297

[6] Chen Y, Xie H, Ma M, Kang Y, Gao X, Shi L, Cao Y, Gao X, Fan H, Wen M, Zeng J, Ghosh S, Zhang X, Zhang C, Lin Q, Rajmohan S, Zhang D, Xu T (2023) Automatic root cause analysis via large language models for cloud incidents. arXiv:2305.15778.

[7] Saha A, Hoi SCH (2022) Mining root cause knowledge from cloud service incident investigations for AIOps. arXiv:2204.11598.

[8] Zhang Y, Guan Z, Qian H, Xu L, Liu H, Wen Q, Sun L, Jiang J, Fan L, Ke M (2021) CloudRCA: a root cause analysis framework for cloud computing platforms. arXiv:2111.03753.

[9] Xu J, Xu Z, Shi B (2022) Deep reinforcement learning based resource allocation strategy in cloud-edge computing system. Frontiers in Bioengineering and Biotechnology 10:908056. https://doi.org/10.3389/fbioe.2022.908056

[10] Gari Y, Monge DA, Pacini E, Mateos C, Garino CG (2021) Reinforcement learning-based application autoscaling in the cloud: a survey. Engineering Applications of Artificial Intelligence 102:104288. https://doi.org/10.1016/j.engappai.2021.104288

[11] Xue S, Qu C, Shi X, Liao C, Zhu S, Tan X, Ma L, Wang S, Wang S, Hu Y, Lei L, Zheng Y, Li J, Zhang J (2022) A meta reinforcement learning approach for predictive autoscaling in the cloud. arXiv:2205.15795.

[12] Fettes Q, Karanth A, Bunescu R, Beckwith B, Subramoney S (2023) Reclaimer: a reinforcement learning approach to dynamic resource allocation for cloud microservices. arXiv:2304.07941.

[13] Abdel Khaleq A, Ra I (2023) Intelligent microservices autoscaling module using reinforcement learning. Cluster Computing

26:2789-2800. https://doi.org/10.1007/s10586-023-03999-8

[14] Nobre J, Pires EJS, Reis A (2023) Anomaly detection in microservice-based systems. Applied Sciences 13(13):7891. https://doi.org/10.3390/app13137891

[15] Kohyarnejadfard I, Aloise D, Azhari SV, Dagenais MR (2022) Anomaly detection in microservice environments using distributed tracing data analysis and NLP. Journal of Cloud Computing 11:25. https://doi.org/10.1186/s13677-022-00296-4

[16] Notaro P, Cardoso J, Gerndt M (2020) A systematic mapping study in AIOps. arXiv:2012.09108.

[17] Jiang Z, Li T, Zhang Z, Ge J, You J, Li L (2021) A survey on log research of AIOps: methods and trends. Mobile Networks and Applications 26:2353-2364. https://doi.org/10.1007/s11036-021-01832-3

[18] Zolanvari M, Ghubaish A, Jain R (2022) ADDAI: anomaly detection using distributed AI. arXiv:2205.01231.

[19] Fernando D, Rodriguez MA, Buyya R (2024) iAnomaly: a toolkit for generating performance anomaly datasets in edge-cloud integrated computing environments. arXiv:2411.02868.

[20] National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

[21] National Institute of Standards and Technology (2024) The NIST Cybersecurity Framework 2.0. National Institute of Standards and Technology, Gaithersburg.

[22] European Union Agency for Cybersecurity (2024) ENISA Threat Landscape 2024. ENISA, Athens.

[23] OWASP Foundation (2025) OWASP Top 10 for Large Language Model Applications 2025. OWASP, online publication.

[24] International Organization for Standardization and International Electrotechnical Commission (2022) ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection - Information security management systems - Requirements. ISO, Geneva.

[25] International Organization for Standardization and International Electrotechnical Commission (2023) ISO/IEC 42001:2023 Information technology - Artificial intelligence - Management system. ISO, Geneva.

[26] Amazon Web Services (2024) AWS Well-Architected Framework: Reliability Pillar. Amazon Web Services, Seattle.

[27] Microsoft (2024) Azure Well-Architected Framework: Reliability. Microsoft, Redmond.

[28] Google Cloud (2023) Disaster recovery planning guide. Google Cloud Architecture Center, Mountain View.

[29] Cybersecurity and Infrastructure Security Agency (2023) Cross-Sector Cybersecurity Performance Goals. CISA, Washington, DC.

[30] International Organization for Standardization (2022) ISO 22361:2022 Security and resilience - Crisis management - Guidelines. ISO, Geneva.