Business Continuity e Disaster Recovery nell'era dell'Artificial Intelligence

Un approccio dual-use alla resilienza delle infrastrutture critiche

Abstract

L’adozione dell’artificial intelligence da parte degli operatori di infrastrutture critiche ridefinisce la disciplina della business continuity e del disaster recovery. Il lavoro esamina il fenomeno in prospettiva dual-use. Sul versante abilitante, le capability di predictive analytics, anomaly detection e AIOps, l’impiego dei large language model nell’incident response, il self-healing, la difesa dei backup e il digital twin sono mappati sulle fasi del ciclo di continuità. Sul versante di rischio, l’analisi distingue le minacce esogene potenziate dall’AI, le fragilità endogene dei modelli e il rischio sistemico di concentrazione, assumendo l’outage CrowdStrike del luglio 2024 come caso di studio archetipico. La rassegna, condotta su letteratura peer-reviewed del periodo 2021–2026, fonti istituzionali e standard tecnici, mostra che la metodologia consolidata del business continuity management non va sostituita ma estesa: business impact analysis a doppia dimensione, model recovery time e punto di ripristino fidato, MLOps riletto come infrastruttura di recovery, modalità operative degradate e testing convergente. Il quadro regolatorio europeo, dall’AI Act a NIS2, CER e DORA, converge sui medesimi presidi ed eleva per la prima volta backup e fail-safe a requisito di prodotto dei sistemi AI. Il lavoro propone infine un framework operativo di AI-resilient business continuity, articolato in principi di progettazione, modello a tre livelli, scala di maturità e indicatori.

Parole chiave: *business continuity; disaster recovery; artificial intelligence; infrastrutture critiche; resilienza operativa; AI Act; adversarial machine learning.*

Sommario

1. Introduzione

1.1 Contesto e motivazione

Le infrastrutture critiche contemporanee, dalle reti energetiche ai sistemi di trasporto, dai presidi sanitari ai servizi idrici e alle infrastrutture digitali, dipendono in misura crescente da catene tecnologiche complesse, nelle quali la distinzione tradizionale tra *information technology* e *operational technology* si è progressivamente assottigliata. In questo contesto la capacità di un'organizzazione di continuare a erogare prodotti e servizi entro tempi e livelli predefiniti a fronte di un evento avverso, che lo standard ISO 22301 formalizza nella nozione di *business continuity*,[1] ha assunto il rango di requisito esistenziale prima ancora che di obbligo normativo. Il *disaster recovery*, inteso come l'insieme delle misure tecnologiche e organizzative destinate al ripristino dei sistemi informativi e dei dati dopo un'interruzione, ne costituisce la componente attuativa sul piano ICT.

La pressione esercitata dal *threat landscape* su questi assetti è documentata con regolarità dall'Agenzia dell'Unione europea per la cibersicurezza. L'edizione 2025 del rapporto ENISA Threat Landscape analizza 4.875 incidenti registrati tra il luglio 2024 e il giugno 2025 e rileva che i soggetti essenziali, come definiti dalla direttiva NIS2, rappresentano il 53,7% del totale degli eventi censiti; il *phishing* si conferma il vettore di accesso iniziale dominante, con circa il 60% dei casi, mentre i sistemi di *operational technology* esposti su Internet figurano tra gli obiettivi a più alto valore per l'intero spettro degli attori di minaccia.[2]

In questo scenario l'*artificial intelligence* entra con una duplice valenza, che costituisce l'oggetto del presente lavoro. Sul primo versante gli operatori la adottano come tecnologia di difesa e di ottimizzazione: le tecniche di *machine learning* alimentano la *predictive maintenance* degli asset industriali, l'*anomaly detection* sul traffico di rete e sui log applicativi, l'automazione dei processi di *incident response*. La letteratura più recente documenta inoltre l'ingresso dei *large language model* nei Security Operations Center, dove automatizzano l'analisi dei log, accelerano il triage degli alert e supportano gli analisti nelle fasi di investigazione e di reporting.[3] Applicate al ciclo di vita della continuità operativa, queste capacità promettono di comprimere i tempi di rilevazione e di ripristino, dunque di agire direttamente sulle metriche che il *business continuity management* presidia.

Sul versante opposto la medesima tecnologia amplia la superficie di attacco e introduce dipendenze inedite. Lo stesso rapporto ENISA segnala l'impiego sistematico di *large language model* per potenziare *phishing* e *social engineering*, al punto che a inizio 2025 le campagne di

[1] International Organization for Standardization, «ISO 22301:2019 – Security and resilience – Business continuity management systems – Requirements», Ginevra, 2019.

[2] European Union Agency for Cybersecurity (ENISA), «ENISA Threat Landscape 2025», ottobre 2025, disponibile su: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025.

[3] A. Habibzadeh, F. Feyzi, R. Ebrahimi Atani, «Large Language Models for Security Operations Centers: A Comprehensive Survey», arXiv:2509.10858, 2025, DOI: 10.48550/arXiv.2509.10858.

phishing supportate da AI rappresentavano, secondo le stime ivi riportate, oltre l'80% dell'attività di social engineering osservata a livello globale, e registra una crescita degli attacchi alla *AI supply chain*.[4] Sul piano tassonomico il National Institute of Standards and Technology ha consolidato nel rapporto AI 100-2e2025 la classificazione degli attacchi di *adversarial machine learning*: *evasion*, *poisoning* e attacchi alla privacy per i sistemi predittivi; *supply chain attack*, *prompt injection* diretta e indiretta e compromissione degli AI agent per i sistemi generativi.[5] A queste minacce deliberate si affiancano fragilità intrinseche dei sistemi di machine learning, quali il *model drift*, le *hallucination* dei modelli generativi e l'*automation bias* degli operatori umani, che possono degradare la qualità delle decisioni proprio nei momenti in cui la continuità del servizio è più esposta.

Che la fragilità delle catene digitali automatizzate e fortemente concentrate non sia un'ipotesi teorica lo ha mostrato l'incidente CrowdStrike del 19 luglio 2024, quando un aggiornamento difettoso di un componente di *endpoint protection* ha provocato il crash di circa 8,5 milioni di dispositivi Windows,[6] con la cancellazione di voli commerciali e l'interruzione di servizi ospedalieri; il Government Accountability Office statunitense lo annovera tra i più estesi IT outage della storia.[7] Le lezioni formalizzate dalla Financial Conduct Authority britannica insistono sull'identificazione dei *single point of failure*, sul change management dei fornitori con accesso profondo ai sistemi e sull'adozione di rilasci progressivi degli aggiornamenti.[8] L'episodio non ha avuto origine da un sistema di AI; costituisce nondimeno l'archetipo della classe di rischio che l'adozione su larga scala di componenti automatizzate, distribuite da pochi fornitori globali, tende ad amplificare, e che i capitoli successivi esaminano nella declinazione specifica dell'artificial intelligence.

Il legislatore europeo ha reagito a questo quadro con un corpus normativo che eleva la resilienza a obbligo giuridico presidiato da sanzioni: la direttiva NIS2 in materia di cibersicurezza dei soggetti essenziali e importanti,[9] la direttiva CER sulla resilienza dei soggetti critici,[10] il regolamento DORA sulla resilienza operativa digitale del settore finanziario,

[4]ENISA, «ENISA Threat Landscape 2025», cit.

[5]National Institute of Standards and Technology, «Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations», NIST AI 100-2e2025, marzo 2025, disponibile su: https://csrc.nist.gov/pubs/ai/100/2/e2025/final.

[6]Microsoft, «Helping our customers through the CrowdStrike outage», The Official Microsoft Blog, 20 luglio 2024.

[7]U.S. Government Accountability Office, «Cyber Resiliency: CrowdStrike Outage Highlights Challenges», GAO-24-107733, settembre 2024, disponibile su: https://www.gao.gov/products/gao-24-107733.

[8]Financial Conduct Authority, «CrowdStrike outage: lessons for operational resilience», 31 ottobre 2024, disponibile su: https://www.fca.org.uk/firms/operational-resilience/crowdstrike-outage-lessons-operational-resilience.

[9]Direttiva (UE) 2022/2555 del Parlamento europeo e del Consiglio, del 14 dicembre 2022 (c.d. direttiva NIS2), GU L 333 del 27.12.2022; recepita in Italia con il d.lgs. 4 settembre 2024, n. 138.

[10]Direttiva (UE) 2022/2557 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativa alla resilienza dei soggetti critici (c.d. direttiva CER), GU L 333 del 27.12.2022; recepita in Italia con il d.lgs. 4 settembre 2024, n. 134.

applicabile dal 17 gennaio 2025,[11] e l'AI Act, che impone ai sistemi ad alto rischio requisiti di accuratezza, robustezza e cybersecurity lungo l'intero ciclo di vita.[12] La convergenza di questi strumenti produce un effetto che la prassi non ha ancora pienamente metabolizzato: la continuità operativa non riguarda più soltanto i processi supportati dall'AI, ma i sistemi di AI in quanto tali, che divengono asset critici la cui indisponibilità, compromissione o deriva qualitativa deve essere prevenuta, gestita e, in determinate circostanze, notificata alle autorità.

La letteratura scientifica ha sinora affrontato i due versanti in filoni largamente separati: gli studi sull'impiego dell'AI a supporto della resilienza operativa procedono in parallelo rispetto a quelli sulla sicurezza e sull'affidabilità dei sistemi di machine learning, mentre i framework di business continuity management continuano per lo più a trattare l'AI come tecnologia abilitante anziché come oggetto di protezione. Il presente lavoro muove dalla convinzione che tale separazione non sia più sostenibile per gli operatori di infrastrutture critiche e adotta una prospettiva *dual-use*, sintetizzata nella Figura 1.1: *AI for resilience* e *resilience of AI* come componenti inscindibili di una medesima strategia di continuità.

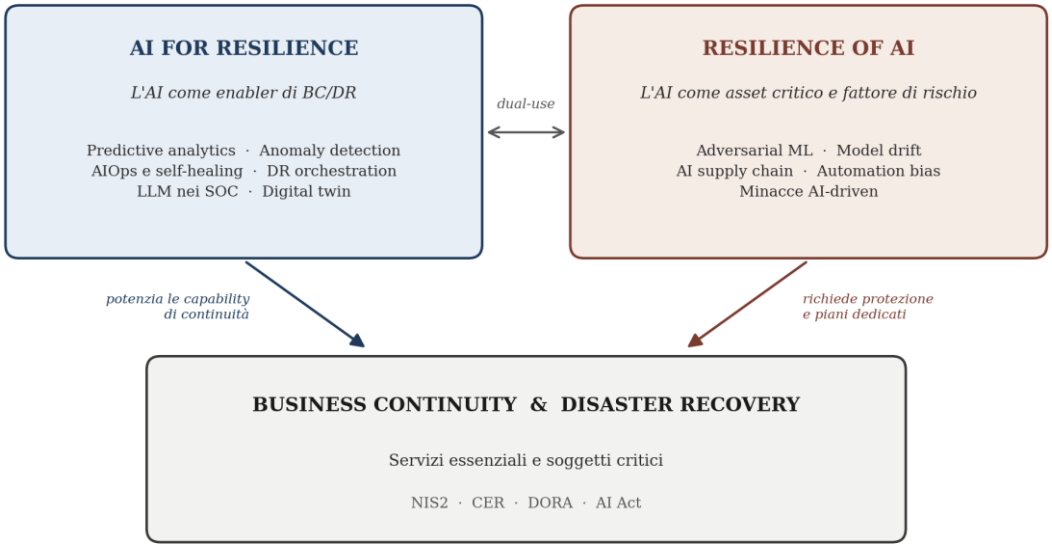

Figura 1.1 – La prospettiva dual-use adottata nel presente lavoro: l'artificial intelligence come enabler delle capability di business continuity e disaster recovery e, al contempo, come asset critico e fattore di rischio da presidiare.

1.2 Obiettivi e research question

[11] Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio, del 14 dicembre 2022 (c.d. regolamento DORA), GU L 333 del 27.12.2022.

[12] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024 (c.d. AI Act), GU L del 12.7.2024, art. 15.

Coerentemente con la prospettiva delineata, il lavoro si propone di sistematizzare la letteratura recente e di derivarne indicazioni operative per gli operatori di infrastrutture critiche. Gli obiettivi sono formalizzati in quattro *research question*.

RQ1. In che modo le tecniche di artificial intelligence possono potenziare le capability di business continuity e disaster recovery degli operatori di infrastrutture critiche, e con quale grado di maturità documentato in letteratura?

RQ2. Quali nuove classi di rischio per la continuità operativa derivano dall'adozione dell'AI, considerata sia come vettore di minaccia sia come fonte di fragilità intrinseche e sistemiche?

RQ3. Come devono evolvere le metodologie consolidate del business continuity management, dalla business impact analysis alla definizione di RTO e RPO fino al testing, per trattare modelli, dataset e pipeline di machine learning come asset critici?

RQ4. In che misura il quadro regolatorio europeo (AI Act, NIS2, CER, DORA) e gli standard volontari definiscono requisiti di resilienza applicabili ai sistemi di AI, e quali lacune permangono?

Alle prime due domande rispondono, rispettivamente, i capitoli 4 e 5; alla terza è dedicato il capitolo 6; la quarta trova risposta nel capitolo 7. Il capitolo 8 ricompone i risultati in un framework operativo di *AI-resilient business continuity*.

1.3 Struttura del capitolo

Il capitolo 2 richiama i fondamenti concettuali e metrici del business continuity management e del disaster recovery, insieme agli standard di riferimento, e ne discute i limiti rispetto al threat landscape corrente. Il capitolo 3 espone l'approccio metodologico seguito nella selezione e nell'analisi delle fonti. I capitoli 4 e 5 sviluppano i due versanti dell'analisi dual-use, dedicati rispettivamente all'AI come enabler e all'AI come fattore di rischio, quest'ultimo corredato dall'esame del caso CrowdStrike. Il capitolo 6 opera la saldatura concettuale tra i due versanti, riconducendo modelli, dataset e pipeline di machine learning alle categorie della business impact analysis. Il capitolo 7 analizza il quadro di governance e compliance, mentre il capitolo 8 propone il framework di sintesi. Chiudono il lavoro le sfide aperte e le direzioni di ricerca (capitolo 9) e le conclusioni (capitolo 10).

2. Background teorico e normativo

2.1 Business continuity e disaster recovery: definizioni e metriche

Nella sistemazione offerta dallo standard ISO 22301, la *business continuity* è la capacità di un'organizzazione di continuare a erogare prodotti e servizi entro tempi accettabili e a capacità predefinita nel corso di un'interruzione; il *business continuity management* è il processo gestionale olistico che identifica le minacce potenziali e i relativi impatti e che costruisce la capacità di risposta dell'organizzazione.[13] La disciplina assume la forma di un sistema di gestione conforme alla struttura armonizzata delle norme ISO, articolato su contesto, leadership, pianificazione, supporto, attività operative, valutazione delle prestazioni e miglioramento continuo; il nucleo operativo risiede nei requisiti sulle attività, dove trovano collocazione business impact analysis e risk assessment, strategie e soluzioni di continuità, piani e procedure, programma di esercitazioni.

La *business impact analysis* costituisce il fondamento analitico dell'intero impianto: identifica i prodotti e i servizi prioritari, ne ricostruisce processi, risorse e dipendenze, interne e di fornitura, e quantifica la crescita degli impatti dell'interruzione nel tempo, ricavandone i requisiti di continuità e di ripristino.[14] Da questa analisi discendono le metriche che governano la progettazione delle soluzioni. Il *maximum tolerable period of disruption* (MTPD, altrimenti detto *maximum acceptable outage*) fissa l'orizzonte temporale oltre il quale gli impatti divengono inaccettabili; il *recovery time objective* (RTO) stabilisce il tempo obiettivo di ripristino di attività e risorse, per costruzione inferiore all'MTPD; il *recovery point objective* (RPO) esprime la massima perdita di dati tollerabile come distanza temporale dall'ultimo punto di ripristino utile, determinando frequenza dei backup e strategie di replica; il *minimum business continuity objective* (MBCO) definisce il livello minimo di servizio accettabile durante l'interruzione, riferimento per le modalità operative degradate.[15] La Tabella 2.1 riepiloga costrutti e riferimenti.

Tabella 2.1 – Metriche fondamentali del business continuity management e principali riferimenti normativi.

Metrica / costrutto	Definizione operativa	Riferimenti principali
Business impact analysis (BIA)	Analizza nel tempo le conseguenze di un'interruzione sulle attività prioritarie, ne ricostruisce risorse e dipendenze e ne deriva i requisiti di continuità e ripristino.	*ISO/TS 22317; ISO 22301, §8.2*
MTPD / MAO	Maximum tolerable period of disruption: intervallo massimo oltre il quale gli impatti dell'indisponibilità di un prodotto o servizio divengono inaccettabili.	*ISO 22300; ISO 22313*

[13] International Organization for Standardization, «ISO 22301:2019», cit.

[14] International Organization for Standardization, «ISO/TS 22317:2021 – Security and resilience – Business continuity management systems – Guidelines for business impact analysis», Ginevra, 2021.

[15] International Organization for Standardization, «ISO 22300:2021 – Security and resilience – Vocabulary», Ginevra, 2021.

Metrica / costrutto	Definizione operativa	Riferimenti principali
RTO	Recovery time objective: tempo obiettivo entro cui ripristinare attività, sistemi o risorse dopo l'incidente; per costruzione RTO ≤ MTPD.	*ISO 22300; NIST SP 800-34*
RPO	Recovery point objective: massima perdita di dati tollerabile, espressa come distanza temporale dall'ultimo punto di ripristino utile; governa frequenza di backup e strategie di replica.	*ISO 22300; NIST SP 800-34*
MBCO	Minimum business continuity objective: livello minimo accettabile di erogazione durante l'interruzione; riferimento per le modalità operative degradate.	*ISO 22300; ISO/IEC 27031*

Il *disaster recovery* occupa, entro questo perimetro, lo spazio della continuità tecnologica: piani, architetture e procedure per il ripristino di sistemi informativi, dati e infrastrutture ICT dopo un'interruzione. La NIST SP 800-34 lo colloca in una famiglia coordinata di piani, dal business continuity plan al disaster recovery plan fino all'information system contingency plan e alla comunicazione di crisi, e ne scandisce il ciclo in sette passi: definizione della policy, business impact analysis, identificazione dei controlli preventivi, elaborazione delle strategie di contingency, sviluppo del piano, attività di testing, training ed exercises, manutenzione.[16] Nelle architetture correnti la funzione si concretizza in siti alternativi, replica sincrona o asincrona, backup immutabili e procedure di failover orchestrate, il cui dimensionamento risponde direttamente a RTO e RPO.

2.2 Gli standard di riferimento: ISO 22301 e NIST SP 800-34

I due riferimenti richiamati assolvono funzioni complementari. ISO 22301:2019 è uno standard di requisiti, certificabile, che disciplina il sistema di gestione nel suo complesso ed è affiancato dalla guida applicativa ISO 22313:2020;[17] il suo baricentro è organizzativo e strategico, e la dimensione tecnologica vi compare come una delle risorse da presidiare. La NIST SP 800-34, concepita per i sistemi informativi federali statunitensi ma divenuta riferimento di fatto anche oltre quel perimetro, opera invece sul piano della pianificazione tecnica di contingency e fornisce il lessico operativo con cui i team ICT dimensionano soluzioni di ripristino e priorità di recovery. La lettura congiunta dei due strumenti restituisce il modello canonico: la BIA traduce gli obiettivi di servizio in RTO e RPO, e questi ultimi guidano investimenti e architetture di disaster recovery.

Il raccordo tra i due piani è stato di recente rafforzato dalla revisione di ISO/IEC 27031, pubblicata nella seconda edizione a maggio 2025 dopo quattordici anni dalla prima: la norma

[16] National Institute of Standards and Technology, «Contingency Planning Guide for Federal Information Systems», NIST Special Publication 800-34 Rev. 1, maggio 2010.

[17] International Organization for Standardization, «ISO 22313:2020 – Security and resilience – Business continuity management systems – Guidance on the use of ISO 22301», Ginevra, 2020.

riposiziona la *ICT readiness for business continuity* come componente strategica della resilienza aziendale, formalizza governance e responsabilità dedicate, integra esplicitamente le dipendenze da cloud e fornitori terzi e mantiene l'aggancio diretto alle metriche MBCO, RTO e RPO, estendendo l'applicabilità delle proprie indicazioni anche a condizioni di attacco attivo, quali campagne ransomware e advanced persistent threat.[18]

Nessuno di questi strumenti, tuttavia, tratta i sistemi di artificial intelligence come categoria autonoma di asset: modelli, dataset di training, feature store e pipeline restano assimilati genericamente alle risorse ICT, senza indicazioni specifiche su come stimarne la criticità, definirne gli obiettivi di ripristino o verificarne la recuperabilità. È un'assunzione che il capitolo 6 discuterà criticamente, alla luce delle peculiarità tecniche che distinguono il ripristino di un modello da quello di un tradizionale sistema transazionale.

2.3 Il threat landscape delle infrastrutture critiche e i limiti dell'approccio tradizionale

Il contesto di minaccia entro cui questi strumenti operano è mutato più rapidamente della loro evoluzione. Il quadro ENISA per il periodo 2024–2025 descrive un ecosistema in cui il ransomware resta la minaccia a maggiore impatto sulle organizzazioni europee, gli attacchi DDoS di matrice hacktivista dominano per volume, le vulnerabilità di nuova pubblicazione vengono sfruttate con crescente rapidità e i sistemi di operational technology raggiungibili da Internet costituiscono bersagli ad alto valore; l'intreccio delle dipendenze digitali fa sì che la perturbazione di un singolo anello si propaghi lungo l'intera supply chain dei servizi essenziali.[19]

Per la disciplina del disaster recovery il ransomware ha un effetto strutturale, perché ne mette in discussione le assunzioni fondative. Le operazioni a doppia estorsione colpiscono deliberatamente le copie di backup e gli strumenti di ripristino, cosicché il recovery cessa di essere un problema di sola disponibilità e diventa un problema di integrità: prima di ripristinare occorre accertare che i punti di ripristino non siano essi stessi compromessi e bonificare l'ambiente, con tempi effettivi che tendono a eccedere gli RTO dichiarati nei piani. Non a caso la revisione 2025 di ISO/IEC 27031 estende esplicitamente la readiness ICT alle condizioni di attacco attivo.[20]

L'incidente CrowdStrike ha aggiunto a questo quadro la dimensione del rischio di concentrazione. Nel trarne le lezioni, la Financial Conduct Authority ha richiamato le imprese a identificare i single point of failure nei propri stack tecnologici, a presidiare il change management dei fornitori con accesso profondo ai sistemi, a testare gli aggiornamenti prima

[18] International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 27031:2025 – Cybersecurity – Information and communication technology readiness for business continuity», 2ª ed., maggio 2025.

[19] ENISA, «ENISA Threat Landscape 2025», cit.

[20] International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 27031:2025», cit.

del rilascio adottando distribuzioni progressive e a includere le terze parti negli scenari di test di resilienza, rilevando peraltro come le problematiche riconducibili a terze parti fossero già risultate la principale causa degli incidenti operativi segnalati nel biennio precedente.[21]

Se ne ricava una diagnosi dei limiti dell'approccio tradizionale che prescinde dall'artificial intelligence e che l'artificial intelligence, come si vedrà, al tempo stesso mitiga e aggrava. La pianificazione classica presuppone minacce a evoluzione lenta rispetto ai cicli annuali di revisione ed esercitazione dei piani; presuppone perimetri di analisi centrati su processi e sistemi convenzionali, con scarsa presa sugli asset algoritmici e sui servizi cognitivi acquisiti da terzi; presuppone, infine, tempi di reazione commisurati alla decisione umana, mentre gli attacchi automatizzati e i guasti a propagazione istantanea comprimono la finestra utile d'intervento sotto la soglia della gestione manuale. Da qui la domanda che i capitoli seguenti affrontano dai due lati della prospettiva dual-use: se e in che misura l'AI possa restituire ai difensori il vantaggio della velocità (capitolo 4), e a quale prezzo in termini di nuove fragilità (capitolo 5).

[21] Financial Conduct Authority, «CrowdStrike outage: lessons for operational resilience», cit.

3. Approccio metodologico

Il presente lavoro adotta la forma della rassegna critica della letteratura, orientata da un impianto analitico definito a priori piuttosto che da un protocollo di revisione sistematica in senso stretto: la natura trasversale del tema, che interseca ingegneria dell'affidabilità, sicurezza informatica, machine learning e regolazione, rende più produttiva una selezione ragionata delle fonti rispetto all'interrogazione esaustiva di singole basi dati. Restano ferme, tuttavia, regole esplicite di inclusione e di verifica.

Il corpus si compone di quattro classi di fonti. La letteratura scientifica peer-reviewed pubblicata tra il 2021 e il 2026, reperita principalmente attraverso IEEE Xplore, ACM Digital Library, ScienceDirect e SpringerLink, costituisce l'ossatura dell'analisi; i preprint depositati su arXiv sono ammessi soltanto se ampiamente citati dalla letteratura successiva o se destinati a sedi con revisione paritaria, e sono indicati come tali nelle note. Le pubblicazioni istituzionali di agenzie e autorità di vigilanza (ENISA, NIST, GAO, FCA) e gli standard ISO/IEC forniscono l'ancoraggio tecnico-normativo; gli atti legislativi dell'Unione europea sono citati nella versione pubblicata in Gazzetta ufficiale. Per il background metodologico sono ammessi riferimenti anteriori quando fondativi, come nel caso della NIST SP 800-34. Sono escluse le fonti promozionali prive di riscontri indipendenti; ciascun riferimento è stato verificato sulla fonte primaria.

L'analisi procede lungo la griglia dual-use introdotta nel capitolo 1. Per il versante abilitante, gli use case sono mappati sulle fasi del ciclo di vita del business continuity management e valutati rispetto alle metriche su cui incidono (Tabella 4.1), distinguendo, ove le fonti lo consentono, il grado di maturità tra evidenza di ricerca, sperimentazione pilota e impiego in produzione. Per il versante di rischio, la classificazione distingue minacce esogene potenziate dall'AI, fragilità endogene dei sistemi di machine learning e rischi sistemici di concentrazione, ricondotti agli impatti sulla continuità operativa (Tabella 5.1). Il caso CrowdStrike è trattato come caso di studio strumentale, selezionato per la disponibilità di fonti primarie di matrice regolatoria e per il valore archetipico rispetto alla classe di rischio esaminata.

4. AI come enabler di business continuity e disaster recovery

La letteratura sull'impiego dell'artificial intelligence a supporto delle operations ha raggiunto negli ultimi anni una massa critica sufficiente a consentirne una lettura organica. Una survey di riferimento sul dominio AIOps articola il campo in quattro compiti fondamentali, incident detection, failure prediction, root cause analysis e automated actions,[22] che corrispondono in modo quasi speculare alle esigenze del ciclo di continuità: rilevare prima, prevedere, comprendere, agire. Questo capitolo ripercorre le sei famiglie di capability più rilevanti per gli operatori di infrastrutture critiche; la Tabella 4.1 le colloca sulle fasi del ciclo di vita del business continuity management e ne indica le metriche interessate, mentre la Figura 4.1, in chiusura di capitolo, ne ricompone l'architettura d'insieme.

Tabella 4.1 – Tassonomia degli use case dell'artificial intelligence mappati sul ciclo di vita del business continuity management.

Fase del ciclo BCM	Use case rappresentativi	Tecniche prevalenti	Contributo alle metriche
Analisi e pianificazione (BIA, risk assessment)	Estrazione di dipendenze e scenari dalla documentazione; prioritizzazione dei rischi	NLP, knowledge graph	*Copertura e aggiornamento della BIA*
Prevenzione	Predictive maintenance degli asset; previsione di guasti e di saturazioni di capacità	Modelli su serie temporali, deep learning (stima RUL)	*Riduzione delle interruzioni non pianificate*
Rilevazione	Anomaly detection su log, rete e telemetria OT; failure prediction	Autoencoder, modelli sequenziali, isolation forest	*MTTD*
Risposta	Triage e correlazione degli alert; sintesi degli incidenti; incident response planning assistito	LLM, SOAR, root cause analysis	*MTTR*
Ripristino	Orchestrazione del failover; self-healing; selezione dei restore point integri	Automazione a runbook, controllo a feedback	*RTO e RPO effettivi*
Esercitazione e miglioramento	Simulazione di scenari su digital twin; stress test e prove di ripristino continue	Digital twin, simulazione	*Frequenza e realismo dei test*

4.1 Predictive analytics ed early warning

La manutenzione predittiva costituisce l'applicazione più matura del machine learning alla prevenzione delle interruzioni nei settori a forte intensità di asset fisici, dall'energia ai trasporti al ciclo idrico. Modelli addestrati su dati di condizione, quali vibrazioni, temperature, assorbimenti elettrici e telemetrie di processo, riconoscono i precursori del guasto e stimano la remaining useful life dei componenti; la survey di Serradilla e colleghi sistematizza le architetture di deep learning applicabili alle diverse fasi della predictive maintenance e ne

[22] Q. Cheng, D. Sahoo, A. Saha, W. Yang, C. Liu, G. Woo, M. Singh, S. Savarese, S. C. H. Hoi, «AI for IT Operations (AIOps) on Cloud Platforms: Reviews, Opportunities and Challenges», arXiv:2304.04661, 2023, DOI: 10.48550/arXiv.2304.04661.

valuta la rispondenza ai requisiti industriali, dalla gestione della variabilità dei dati all'adattabilità dei modelli nel tempo.[23]

Sul piano della continuità operativa l'effetto è duplice. La prevenzione riduce la frequenza degli eventi che i piani devono gestire; la previsione della finestra di guasto consente di convertire interruzioni non pianificate in fermi programmati, con impatti negoziati anziché subiti. La logica dell'early warning si estende peraltro oltre l'asset fisico: previsione delle saturazioni di capacità, stress indotto da eventi meteorologici sulle reti di distribuzione, segnali precoci di degrado applicativo. In termini di business impact analysis, queste capability agiscono sulla probabilità dello scenario avverso prima ancora che sulla risposta.

4.2 Anomaly detection e AIOps

L'osservabilità dei sistemi distribuiti produce volumi di log, metriche e trace che eccedono strutturalmente la capacità di analisi manuale, con il corollario ben noto dell'alert fatigue. L'AIOps risponde applicando tecniche di machine learning a questi flussi: per la rilevazione di anomalie nei log, in particolare, gli approcci di deep learning hanno mostrato prestazioni superiori ai metodi convenzionali nel trattare dati non strutturati e formati instabili,[24] mentre correlazione degli alert, failure prediction e root cause analysis completano la catena che dal segnale grezzo conduce alla diagnosi.[25]

Per la continuità operativa il beneficio si misura anzitutto sul mean time to detect: anticipare la rilevazione significa intervenire quando il degrado è ancora contenibile, prima che l'anomalia diventi indisponibilità e attivi formalmente il piano di continuità. La stessa letteratura segnala però costi di esercizio non banali, dalla qualità dei dati di addestramento alla deriva dei modelli, fino ai falsi positivi che erodono la fiducia degli operatori: temi che il capitolo 5 riprende dal lato delle fragilità.

4.3 LLM e generative AI nell'incident response e nei Security Operations Center

L'ingresso dei large language model nei Security Operations Center è lo sviluppo più recente e di più rapida diffusione. La prima survey sistematica dedicata al tema documenta applicazioni lungo l'intero flusso di lavoro: automazione dell'analisi dei log, accelerazione del triage, miglioramento dell'accuratezza della detection, accesso immediato a conoscenza specialistica altrimenti scarsa,[26] in risposta a criticità croniche dei SOC quali i volumi di alert e la carenza di analisti esperti.

---

[23] O. Serradilla, E. Zugasti, J. Rodriguez, U. Zurutuza, «Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects», Applied Intelligence, vol. 52, n. 10, 2022, pp. 10934–10964, DOI: 10.1007/s10489-021-03004-y.

[24] M. Landauer, S. Onder, F. Skopik, M. Wurzenberger, «Deep learning for anomaly detection in log data: A survey», Machine Learning with Applications, vol. 12, 100470, 2023, DOI: 10.1016/j.mlwa.2023.100470.

[25] Q. Cheng, D. Sahoo et al., «AI for IT Operations (AIOps) on Cloud Platforms», cit.

[26] A. Habibzadeh, F. Feyzi, R. Ebrahimi Atani, «Large Language Models for Security Operations Centers», cit.

La frontiera della ricerca riguarda il passaggio dal supporto informativo alla pianificazione operativa. Lavori recenti dimostrano l'impiego di modelli linguistici compatti per la generazione di piani di incident response con tecniche esplicite di contenimento delle hallucination,[27] mentre prototipi di risposta automatizzata end-to-end esplorano l'esecuzione assistita dell'intero ciclo di gestione dell'incidente.[28] A questi usi si affianca, sul piano della crisis communication, la redazione assistita di situation report e comunicazioni verso gli stakeholder, attività che nei piani di continuità assorbe tempo pregiato durante l'emergenza. L'affidabilità resta condizionata dal controllo delle hallucination e dalla resistenza alla prompt injection: per questa ragione l'impiego operativo corrente privilegia configurazioni assistive, con validazione umana delle decisioni.

4.4 Self-healing infrastructure e disaster recovery orchestration

Il passo ulteriore è la chiusura del ciclo, dall'insight all'azione. Le architetture self-healing eseguono runbook di rimedio al verificarsi di condizioni riconosciute: riavvio di servizi, scaling delle risorse, isolamento di nodi compromessi. Su un piano più avanzato, la ricerca sulla tolleranza alle intrusioni ha dimostrato sistemi che mantengono il servizio sotto attacco riconfigurandosi mediante controllo a retroazione su due livelli.[29]

Nel dominio del disaster recovery l'automazione intelligente investe l'orchestrazione del failover, la verifica continua della coerenza tra sito primario e sito di ripristino e l'esecuzione ricorrente di prove di recovery, trasformando il test da evento eccezionale a controllo di routine. Occorre tuttavia distinguere la ricerca dalla pratica corrente: negli ambienti di produzione prevalgono automazioni deterministiche con machine learning in funzione ausiliaria, e le azioni irreversibili o ad alto impatto restano di norma soggette ad approvazione umana, con periodi di esercizio in shadow mode nelle fasi di adozione. È una prudenza coerente con il principio di supervisione che il quadro regolatorio, come si vedrà nel capitolo 7, eleva a obbligo.

4.5 Cyber resilience dei backup: ransomware detection e immutable storage

La letteratura sul ransomware documenta come le operazioni moderne colpiscano deliberatamente le copie di backup e gli strumenti di ripristino, precondizione per massimizzare la pressione estorsiva;[30] il quadro ENISA conferma la centralità della minaccia per le

[27] K. Hammar, T. Alpcan, E. C. Lupu, «Incident Response Planning Using a Lightweight Large Language Model with Reduced Hallucination», in Proceedings of the 33rd Annual Network and Distributed System Security Symposium (NDSS 2026), San Diego, 2026.

[28] X. Lin, J. Zhang, G. Deng, T. Liu, X. Liu, C. Yang, T. Zhang, Q. Guo, R. Chen, «IRCopilot: Automated Incident Response with Large Language Models», arXiv:2505.20945, 2025.

[29] K. Hammar, R. Stadler, «Intrusion Tolerance for Networked Systems through Two-Level Feedback Control», in Proceedings of the 54th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2024), 2024, pp. 338–352.

[30] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, M. K. Khan, «Ransomware: Recent advances, analysis, challenges and future research directions», Computers & Security, vol. 111, 102490, 2021, DOI: 10.1016/j.cose.2021.102490.

organizzazioni europee.[31] La risposta architetturale passa per lo storage immutabile e per le copie isolate; il contributo specifico dell'AI si colloca sul piano della rilevazione e della selezione: identificazione dei pattern di cifratura massiva attraverso variazioni di entropia e tassi anomali di modifica dei file, classificazione dei restore point per individuare l'ultimo punto integro precedente alla compromissione, stima del perimetro impattato a supporto della bonifica.

In termini di metriche, queste capability difendono l'RPO reale, che una compromissione silente dei backup può degradare di settimane rispetto a quello nominale, e comprimono la componente più onerosa dell'RTO negli scenari ransomware, ossia il tempo di accertamento dell'integrità. Qui trova attuazione concreta l'estensione della ICT readiness alle condizioni di attacco attivo sancita dalla revisione 2025 di ISO/IEC 27031.[32]

4.6 Digital twin e simulazione di scenari

Il digital twin, inteso come replica virtuale connessa in modo bidirezionale al sistema fisico, offre alla disciplina della continuità ciò che più le è mancato: un ambiente in cui sollecitare l'infrastruttura senza toccare la produzione. Sulle repliche è possibile condurre analisi what-if su guasti e attacchi, stress test di capacità, validazione preventiva dei piani di ripristino e addestramento dei team di risposta su scenari realistici; la survey di Alcaraz e Lopez ne sistematizza architetture e livelli funzionali, evidenziando al contempo che il twin stesso, in quanto specchio fedele dell'impianto, diventa un obiettivo da proteggere con lo stesso rigore dell'originale.[33]

Per gli operatori di infrastrutture critiche il valore risiede nel superamento del limite diagnosticato nel capitolo 2: l'esercitazione smette di essere l'evento annuale a perimetro ridotto e diventa processo continuo, ripetibile e misurabile, con la possibilità di rigiocare incidenti reali e varianti sintetiche degli stessi. La generazione assistita di scenari mediante modelli linguistici estende ulteriormente il repertorio delle esercitazioni, ferma restando la necessità di validazione esperta dei contenuti generati.

Ricomposte in architettura, le capability descritte disegnano una pipeline coerente: dalle sorgenti di telemetria al livello di intelligenza, dall'orchestrazione delle azioni agli esiti misurabili sulle metriche richiamate nel capitolo 2, sotto una fascia trasversale di supervisione umana e governance (Figura 4.1). La stessa rappresentazione rende però visibile il rovescio della medaglia: ogni blocco aggiunge software, modelli, dati e dipendenze da terzi che entrano a loro volta nel perimetro da proteggere. Ai rischi che ne derivano è dedicato il capitolo 5.

[31] ENISA, «ENISA Threat Landscape 2025», cit.

[32] International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 27031:2025», cit.

[33] C. Alcaraz, J. Lopez, «Digital Twin: A Comprehensive Survey of Security Threats», IEEE Communications Surveys & Tutorials, vol. 24, n. 3, 2022, pp. 1475–1503, DOI: 10.1109/COMST.2022.3171465.

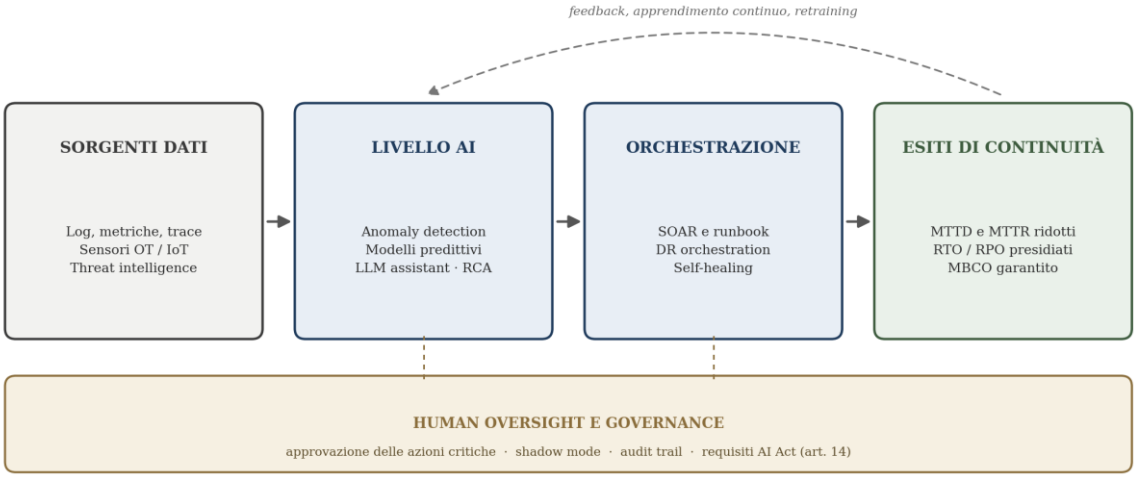

Figura 4.1 – Architettura di riferimento AI-augmented per business continuity e disaster recovery: pipeline dalle sorgenti di telemetria agli esiti di continuità, con supervisione umana trasversale.

5. AI come fattore di rischio per la continuità operativa

Speculare al precedente, questo capitolo esamina il versante di rischio della prospettiva dual-use lungo tre direttrici: le minacce esogene che l'artificial intelligence potenzia nelle mani degli attaccanti, le fragilità endogene dei sistemi di machine learning che gli operatori integrano nei propri processi, e i rischi sistemici che derivano dalla struttura concentrata dell'ecosistema AI. Chiude il capitolo l'esame del caso CrowdStrike, che di questa terza direttrice costituisce l'archetipo documentato. La Tabella 5.1 anticipa la mappatura complessiva delle classi di rischio sugli impatti di continuità e sui relativi presidi.

Tabella 5.1 – Mappatura delle classi di rischio AI sugli impatti di business continuity e disaster recovery.

Classe di rischio	Manifestazioni tipiche	Impatto sulla continuità	Presidi principali
Minacce esogene AI-driven	Phishing e vishing sintetici, deepfake audio-video, malware assistito da AI	Accesso iniziale più efficace; frode nei processi decisionali di crisi	*Verifiche out-of-band, procedure anti-impersonation, awareness*
Attacchi adversarial ai modelli	Evasion, poisoning, prompt injection; manipolazione di assistenti LLM e agent	Degrado silente dei controlli automatici proprio durante l'incidente	*Adversarial testing, validazione di input e output, isolamento dei tool*
Fragilità endogene	Model drift, hallucination, automation bias	Decisioni errate a velocità macchina; fiducia mal calibrata degli operatori	*Monitoraggio delle performance, soglie di confidenza, human-in-the-loop*
Rischio sistemico e di concentrazione	Dipendenza da foundation model, API e hyperscaler; compromissione della AI supply chain	Indisponibilità simultanea multi-organizzazione; common mode failure	*Diversificazione, exit strategy, requisiti contrattuali, scenari con terze parti*
Automazione come amplificatore	Update difettosi o malevoli a propagazione globale (caso CrowdStrike)	Outage istantaneo su larga scala con ripristino manuale	*Rilasci progressivi, canary, rollback e recovery indipendenti dallo strumento*

5.1 Minacce AI-driven: social engineering sintetico, deepfake, malware adattivo

Il primo effetto misurabile dell'AI sul threat landscape riguarda l'industrializzazione dell'inganno. Il quadro ENISA documenta l'impiego sistematico di large language model per potenziare le operazioni offensive esistenti, dalla generazione di malspam alla conduzione di campagne di social engineering: secondo le stime riportate nel rapporto, a inizio 2025 le campagne di phishing supportate da AI costituivano oltre l'80% dell'attività di social engineering osservata a livello mondiale, mentre l'emergere di sistemi di AI malevoli dedicati solleva interrogativi sulle capacità offensive future.[34]

Il capitolo più insidioso per la continuità operativa è quello dei media sintetici. La letteratura sulla creazione e sulla rilevazione dei deepfake mostra architetture generative in costante

[34] ENISA, «ENISA Threat Landscape 2025», cit.

progresso e difese che faticano a generalizzare oltre i dataset di addestramento;[35] audio e video sintetici in tempo reale rendono praticabile l'impersonation di dirigenti e fornitori nelle comunicazioni operative. Il punto di contatto con la disciplina di questo lavoro è preciso: i processi di crisi si reggono su comunicazioni fuori dai canali ordinari, telefonate di escalation, approvazioni urgenti, istruzioni di failover, ed è esattamente in queste condizioni di pressione e di deroga che l'impersonation sintetica trova il terreno migliore. I piani di continuità devono perciò incorporare verifiche out-of-band e parole d'ordine procedurali che non transitino sul canale potenzialmente compromesso. Sul fronte del codice, l'assistenza dei modelli generativi comprime i cicli di sviluppo del malware e ne facilita la variazione continua, erodendo l'efficacia delle difese basate su firme.

5.2 Adversarial machine learning: la tassonomia NIST AI 100-2

La seconda direttrice riguarda gli attacchi ai sistemi di machine learning in quanto tali. La tassonomia consolidata dal NIST distingue, per i sistemi predittivi, tre famiglie: gli attacchi di evasion, che manipolano gli input in fase di inferenza per indurre classificazioni errate; il poisoning, che corrompe dati di addestramento o modello per degradarne il comportamento o innestarvi backdoor; gli attacchi alla privacy, che estraggono informazioni sui dati di training o sul modello stesso. Per i sistemi generativi la tassonomia aggiunge gli attacchi alla supply chain, la prompt injection diretta e indiretta, le violazioni d'uso e la sicurezza degli AI agent.[36] La chiave di lettura più utile è quella del ciclo di vita: ogni fase, dalla raccolta dei dati al monitoraggio in esercizio, espone vettori propri, come schematizza la Figura 5.1.

Per la continuità operativa la peculiarità di questi attacchi è il degrado silente. Un sistema convenzionale compromesso tende a manifestare malfunzionamenti; un modello avvelenato continua a rispondere, soltanto peggio, e un assistente vittima di prompt injection esegue azioni formalmente legittime per finalità ostili. Poiché tra i bersagli figurano le stesse capability difensive descritte nel capitolo 4, l'effetto è una perdita di efficacia dei controlli che non attiva alcun allarme: la detection si spegne gradualmente mentre i cruscotti restano verdi. La superficie interessata comprende inoltre componenti che i piani di continuità tradizionali non censiscono (dataset, feature store, model registry); il capitolo 6 li riporterà dentro il perimetro dell'analisi.

[35] Y. Mirsky, W. Lee, «The Creation and Detection of Deepfakes: A Survey», ACM Computing Surveys, vol. 54, n. 1, art. 7, 2021, pp. 1–41, DOI: 10.1145/3425780.

[36] National Institute of Standards and Technology, «Adversarial Machine Learning», NIST AI 100-2e2025, cit.

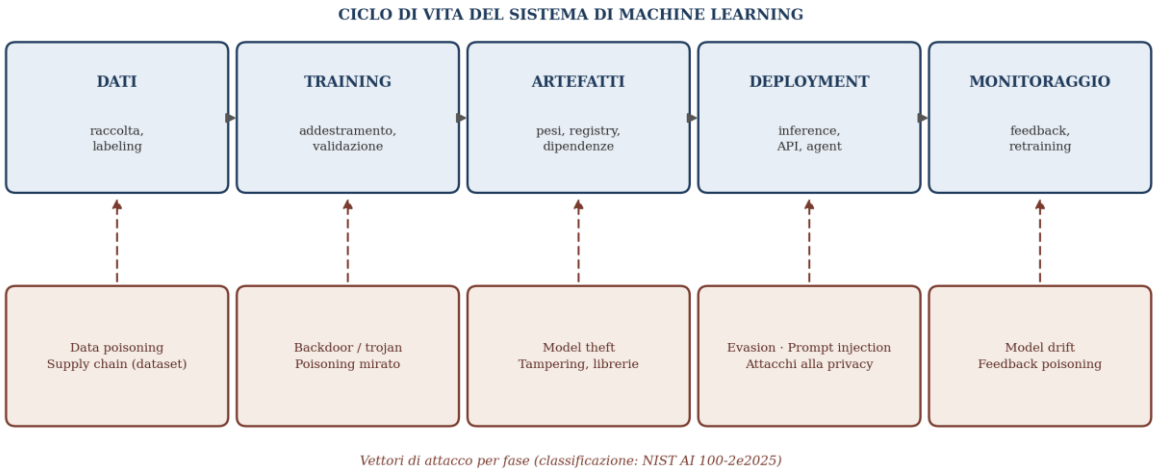

Figura 5.1 – Superficie di attacco del ciclo di vita di un sistema di machine learning, con i vettori principali per fase secondo la classificazione NIST AI 100-2e2025.

5.3 Fragilità intrinseche: model drift, hallucination, automation bias

Il terzo ordine di problemi non richiede alcun avversario. I modelli di machine learning degradano spontaneamente quando la distribuzione dei dati in esercizio si allontana da quella di addestramento: la letteratura sul concept drift ne ha consolidato tassonomia e metodi di rilevazione, mostrando come il fenomeno accompagni fisiologicamente il ciclo di vita dei sistemi predittivi.[37] Nelle infrastrutture critiche le cause di deriva abbondano: stagionalità dei consumi, modifiche impiantistiche, evoluzione dei pattern di traffico. Un detector addestrato sull'assetto di ieri invecchia in silenzio sull'assetto di oggi, e la sua efficacia va trattata come una grandezza da monitorare, non come una proprietà acquisita.

I modelli generativi aggiungono la propensione all'hallucination, ossia la produzione di contenuti plausibili ma non fattuali, di cui la letteratura ha proposto tassonomie e tecniche di mitigazione senza pervenire a soluzioni risolutive;[38] in un contesto di incident response, un'informazione falsa ma verosimile entra nella catena decisionale nel momento di massima pressione temporale. Il fattore umano completa il quadro: la ricerca sull'automation bias documenta come l'affidabilità percepita dell'automazione induca compiacenza e riduca il monitoraggio attivo da parte degli operatori,[39] con l'esito paradossale che i sistemi migliori generano la vigilanza peggiore nei confronti dell'errore raro che prima o poi si verifica. Per la continuità operativa la combinazione è critica: errori a velocità macchina incontrano una supervisione umana mal calibrata.

[37] F. Bayram, B. S. Ahmed, A. Kassler, «From concept drift to model degradation: An overview on performance-aware drift detectors», Knowledge-Based Systems, vol. 245, 108632, 2022, DOI: 10.1016/j.knosys.2022.108632.

[38] L. Huang, W. Yu, W. Ma et al., «A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions», ACM Transactions on Information Systems, vol. 43, n. 2, art. 3155, 2025, pp. 1–55, DOI: 10.1145/3703155.

[39] R. Parasuraman, D. H. Manzey, «Complacency and Bias in Human Use of Automation: An Attentional Integration», Human Factors, vol. 52, n. 3, 2010, pp. 381–410, DOI: 10.1177/0018720810376055.

5.4 Rischio sistemico e di concentrazione: AI supply chain e dipendenza dagli hyperscaler

La quarta direttrice attiene alla struttura del mercato. L'ecosistema dell'AI generativa poggia su pochi foundation model, poche piattaforme cloud e una filiera ristretta di hardware specializzato; le organizzazioni vi accedono per lo più tramite API: indisponibilità, modifiche di comportamento o compromissioni presso il fornitore si propagano quindi istantaneamente a valle, senza che gli utilizzatori dispongano di visibilità o di leve dirette. Il quadro ENISA registra nel frattempo una crescita degli attacchi alla AI supply chain, dai dataset ai repository di modelli alle librerie.[40]

La lettura sistemica più strutturata proviene dal Financial Stability Board, che tra le vulnerabilità con potenziale di rischio sistemico indica le dipendenze da terze parti e la concentrazione dei service provider, le correlazioni di mercato indotte da modelli simili, il rischio cyber e il model risk con le connesse questioni di qualità dei dati e di governance;[41] il successivo esercizio di monitoraggio rileva peraltro un controllo crescente di più anelli della filiera da parte dei grandi provider tecnologici.[42] Benché formulata per il settore finanziario, l'analisi si trasferisce senza attrito alle infrastrutture critiche: modelli simili commettono errori simili, e la dipendenza condivisa dallo stesso fornitore trasforma il guasto individuale in common mode failure. Ne discendono due scenari che i piani di continuità raramente contemplano, l'indisponibilità simultanea multi-organizzazione di un servizio AI esterno e la deriva coordinata di comportamenti automatizzati, che meritano ingresso esplicito nel repertorio delle esercitazioni.

5.5 Caso di studio: l'outage CrowdStrike del luglio 2024

I fatti sono documentati con precisione dalle fonti primarie. Il 19 luglio 2024, alle 04:09 UTC, un aggiornamento di configurazione del sensore Falcon per gli host Microsoft Windows conteneva un difetto logico che mandava in crash i sistemi; la distribuzione fu interrotta alle 05:27 UTC, ma i settantotto minuti di esposizione bastarono a colpire circa 8,5 milioni di dispositivi nel mondo,[43] con la cancellazione di voli commerciali, l'interruzione di servizi ospedalieri e impatti diffusi su altri servizi essenziali; il Government Accountability Office lo classifica tra i più estesi IT outage della storia, rilevando come un errore non doloso abbia riprodotto le dinamiche di compromissione della supply chain già osservate in incidenti di matrice ostile.[44]

---

[40] ENISA, «ENISA Threat Landscape 2025», cit.

[41] Financial Stability Board, «The Financial Stability Implications of Artificial Intelligence», 14 novembre 2024, disponibile su: https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/.

[42] Financial Stability Board, «Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector», ottobre 2025.

[43] Microsoft, «Helping our customers through the CrowdStrike outage», cit.

[44] U.S. Government Accountability Office, «Cyber Resiliency: CrowdStrike Outage Highlights Challenges», cit.

La documentazione di campo restituisce la fisionomia del ripristino. Nel sistema sanitario ECU Health, l'aggiornamento difettoso provocò 8.647 eventi di blue screen, circa il 60% del parco dispositivi, di cui 729 richiesero intervento manuale con avvio in safe mode, chiave di ripristino BitLocker e rimozione fisica del file difettoso, mentre il centro di comando rimase operativo per circa trenta ore;[45] la cartella clinica elettronica rimase disponibile anche grazie alla finestra notturna di basso carico, e il personale di informatica clinica funse da cerniera tra team tecnici e reparti. Il dato strutturale è l'asimmetria delle velocità: il guasto si è propagato a velocità software su scala globale, il ripristino è tornato a velocità umana, dispositivo per dispositivo. Nessuna metrica di RTO concepita per guasti localizzati regge un evento che richiede il contatto fisico con centinaia di macchine simultaneamente.

Le lezioni regolatorie insistono sull'identificazione dei single point of failure, sul testing degli aggiornamenti con rilasci progressivi e sull'inclusione delle terze parti negli scenari di resilienza.[46] Ai fini del presente lavoro conta però la lettura prospettica. Il pattern strutturale dell'incidente, un agente con privilegi profondi installato su una flotta globale e alimentato da aggiornamenti automatici considerati fidati, coincide con il pattern di deployment degli strumenti di sicurezza basati su AI e, in prospettiva, degli AI agent operativi: se un file di configurazione difettoso ha prodotto questi effetti, un aggiornamento di modello difettoso o avvelenato, distribuito con il medesimo meccanismo, avrebbe la stessa cinetica di propagazione. L'episodio consegna al capitolo 6 requisiti precisi: capacità di rollback che non dipendano dallo strumento guasto, distribuzione canary anche per i modelli, procedure di ripristino eseguibili quando l'automazione stessa è la causa dell'indisponibilità.

[45] C. R. Dennis, C. S. Evans, K. Duckworth, M. M. Skinner, J. Hanna, T. Thompson, D. Herring, R. J. Medford, «Resilience in the Face of Disruption: Viewpoint on the CrowdStrike Incident in July 2024», JMIR Medical Informatics, vol. 13, e69958, 2025, DOI: 10.2196/69958.

[46] Financial Conduct Authority, «CrowdStrike outage: lessons for operational resilience», cit.

6. Ripensare BC/DR per i sistemi di artificial intelligence

I due capitoli precedenti convergono su una conclusione operativa: l'AI è al tempo stesso strumento di protezione e oggetto da proteggere, ma gli standard di continuità, come rilevato nel capitolo 2, la trattano ancora come una generica risorsa ICT. Questo capitolo compie la saldatura, riconducendo modelli, dati e pipeline alle categorie metodologiche consolidate del business continuity management e mostrando dove tali categorie richiedono estensione. L'esito non è una disciplina nuova, ma la stessa disciplina applicata a una classe di asset con proprietà di guasto, compromissione e ripristino che non hanno equivalenti nei sistemi transazionali.

6.1 Business impact analysis per gli asset AI

Il punto di partenza resta la logica della business impact analysis: dai prodotti e servizi prioritari si risale ai processi e da questi alle risorse che li sostengono.[47] L'estensione riguarda il censimento delle risorse, che deve includere una classe di asset finora invisibile agli inventari di continuità: gli artefatti dei modelli, i dataset di training e validazione, i feature store e le pipeline di alimentazione, i model registry con le relative configurazioni e, per i sistemi generativi, prompt di sistema e definizioni dei tool, oltre ai servizi AI acquisiti da terzi tramite API. Ma l'estensione più profonda riguarda la domanda stessa dell'analisi. Alla domanda classica, quali processi si arrestano se la risorsa è indisponibile, occorre affiancarne una seconda che i sistemi convenzionali non pongono: quali decisioni peggiorano, e con quali conseguenze, se la risorsa degrada senza arrestarsi. È la traduzione metodologica del degrado silente descritto nel capitolo 5.

La valutazione di criticità assume così due dimensioni, disponibilità e affidabilità, i cui driver principali sono il ruolo del sistema nella catena decisionale, dal supporto informativo all'azione autonoma, la reversibilità delle azioni che esso può intraprendere e la sostituibilità umana della funzione svolta. La Tabella 6.1 organizza il censimento proponendo, per ciascuna classe di asset, le domande specifiche di analisi e le implicazioni sugli obiettivi di ripristino. Un ultimo elemento merita menzione: la mappatura delle dipendenze deve estendersi al fornitore del foundation model e alla relativa filiera, secondo la lettura sistemica del capitolo 5, perché la criticità di un processo interno può risiedere in un anello che l'organizzazione non controlla né osserva.

Tabella 6.1 – Gli asset dei sistemi di artificial intelligence ricondotti alle categorie della business impact analysis.

Asset AI	Funzione nel processo	Domande di BIA specifiche	Implicazioni per gli obiettivi di ripristino
Modello (artefatto, pesi)	Capacità predittiva o decisionale del sistema	Quali decisioni dipendono dal modello e con quale autonomia? Qual è l'impatto di un degrado non rilevato?	*RTO dell'artefatto (rollback) distinto dall'RTO della capability (retraining); versioni fidate nel registry*

[47] International Organization for Standardization, «ISO/TS 22317:2021», cit.

Asset AI	Funzione nel processo	Domande di BIA specifiche	Implicazioni per gli obiettivi di ripristino
Dataset di training e validazione	Base di conoscenza del modello	I dati sono ricostruibili? Esistono snapshot immutabili e lineage? Da quale data potrebbero essere avvelenati?	*RPO come ultimo punto fidato, non ultimo disponibile; conservazione di lungo periodo*
Feature store e pipeline dati	Alimentazione del modello in esercizio	Quali sorgenti alimentano le feature? Cosa accade se una sorgente degrada in modo silenzioso?	*Ripristino coordinato di dati e modello; controlli di qualità in ingresso*
Registry, configurazioni, prompt	Comportamento operativo del sistema	Chi può modificarli? Sono versionati, firmati e verificabili?	*Rollback rapido della configurazione; audit trail delle modifiche*
Servizio AI di terze parti (API)	Capability acquisita all'esterno	Qual è l'RTO contrattuale del fornitore? Esiste un fallback: altro fornitore, modello locale, procedura manuale?	*Exit strategy e prova periodica del fallback; scenario di indisponibilità simultanea*

6.2 RTO e RPO per modelli e dati

Le metriche del capitolo 2 restano valide ma si sdoppiano. Per l'RTO occorre distinguere il ripristino dell'artefatto dal ripristino della capability: ricaricare da un registry una versione precedente del modello è questione di minuti, ma se la versione corrente è compromessa da poisoning e quelle precedenti condividono i dati contaminati, il ritorno a una capability fidata può richiedere un riaddestramento su dati bonificati, con tempi che dipendono dalla disponibilità di dataset versionati, dalla capacità di calcolo mobilitabile e dalla durata della validazione. È dunque utile esplicitare, accanto all'RTO tradizionale, un obiettivo di model recovery time riferito al ritorno in esercizio di una versione fidata, dimensionato sullo scenario peggiore del retraining e non su quello ottimistico del rollback.

L'RPO subisce una torsione analoga. Per i dati convenzionali misura la perdita massima tollerabile; per i sistemi di machine learning deve misurare la distanza dall'ultimo punto di ripristino fidato, che non coincide necessariamente con l'ultimo disponibile: un avvelenamento scoperto oggi ma iniziato mesi fa retrodata il punto utile all'inizio della contaminazione, come già osservato per i backup negli scenari ransomware. La perdita dei dataset ha inoltre una gravità peculiare, perché con essi si perde la possibilità di ricostruire il comportamento appreso e di dimostrarne la provenienza. Al perimetro dell'RPO appartengono infine le configurazioni che determinano il comportamento operativo, dai prompt di sistema alle policy dei guardrail, la cui alterazione non tracciata equivale a una modifica non autorizzata del sistema.

6.3 MLOps e model recovery

La disciplina che rende praticabili questi obiettivi esiste già, benché sia nata per finalità di efficienza e non di continuità. Il corpus di pratiche MLOps, di cui la letteratura ha consolidato definizione e architettura di riferimento, prescrive il versionamento congiunto di codice, dati e modelli, pipeline riproducibili di addestramento e rilascio, model registry e tracciamento del

lineage;[48] riletti nella prospettiva di questo lavoro, tali controlli sono esattamente l'infrastruttura di disaster recovery dei sistemi AI. Un registry versionato è il backup del modello; una pipeline riproducibile è la procedura documentata di ripristino; il lineage dei dati è la precondizione per stabilire quale punto sia fidato. L'organizzazione che dispone di un impianto MLOps maturo possiede, spesso senza saperlo, gran parte della propria capacità di model recovery; quella che ne è priva non può comprarla nel momento dell'emergenza.

Su questa base si costruisce il playbook di ripristino: rollback immediato alla versione fidata con distribuzione controllata, ripristino dei dataset da snapshot immutabili, riaddestramento di emergenza con capacità di calcolo contrattualizzata in anticipo, validazione pre-rilascio che includa controlli funzionali e prove di robustezza essenziali. Per le capability acquisite da terzi il playbook assume la forma dell'exit strategy: fornitore alternativo, modello locale di riserva o procedura manuale, con la prova periodica del fallback come unico test credibile della sua esistenza. Vale infine per i modelli la lezione del capitolo 5 sugli aggiornamenti: rilasci progressivi e finestre canary, perché il meccanismo che distribuisce il modello è lo stesso che può distribuirne il difetto.

6.4 Graceful degradation e fallback manuale

Se il ripristino richiede tempo, la continuità nel frattempo richiede modalità degradate progettate in anticipo, secondo la logica che lo standard cristallizza nel minimum business continuity objective.[49] Per i sistemi AI le modalità tipiche sono tre, in ordine crescente di conservazione della funzione: la sospensione completa con ritorno alle procedure manuali, la modalità conservativa in cui il sistema resta attivo ma con soglie prudenziali e blocchi di default, la sospensione selettiva delle sole azioni autonome con mantenimento del supporto informativo. La scelta va compiuta e documentata prima dell'evento, insieme ai trigger oggettivi che attivano il downgrade, tipicamente le metriche di drift e di confidenza discusse nel capitolo 5, perché deciderla durante l'incidente significa deciderla nelle condizioni peggiori.

Due implicazioni operative sono meno ovvie di quanto sembri. La prima è che la disattivazione deve essere possibile e provata: un interruttore documentato, eseguibile da personale autorizzato, che non dipenda dal sistema che si intende disattivare. La seconda è che il fallback manuale presuppone che la competenza manuale esista ancora: l'automation bias erode la vigilanza, ma l'automazione prolungata erode la capacità stessa di operare senza di essa. La contromisura è organizzativa prima che tecnica: esercitazioni periodiche a sistemi AI spenti, che mantengano viva la procedura manuale e ne misurino i tempi reali, i quali a loro volta alimentano il dimensionamento dell'MBCO.

[48] D. Kreuzberger, N. Kühl, S. Hirschl, «Machine Learning Operations (MLOps): Overview, Definition, and Architecture», IEEE Access, vol. 11, 2023, pp. 31866–31879, DOI: 10.1109/ACCESS.2023.3262138.

[49] International Organization for Standardization, «ISO 22300:2021», cit.

6.5 Testing della resilienza: chaos engineering e adversarial testing

La verifica di questo impianto richiede di estendere ai sistemi AI le due tradizioni di test più mature. La prima è il chaos engineering, la pratica di condurre esperimenti controllati sui sistemi in esercizio per verificare ipotesi esplicite di resilienza,[50] che applicata all'AI significa iniettare deliberatamente le condizioni discusse nei capitoli precedenti: indisponibilità e latenza del fornitore esterno, risposte corrotte o incoerenti del modello, degrado simulato delle metriche di confidenza, per osservare se i trigger di downgrade scattano, se i fallback reggono il carico e se gli operatori riconoscono la transizione. Un esercizio di questo tipo, condotto come game day sugli scenari di AI indisponibile e di AI non fidata, vale più di qualunque attestazione documentale.

La seconda tradizione è il testing avversario, che verifica la tenuta dei modelli rispetto alle famiglie di attacco della tassonomia NIST, dall'evasion alla prompt injection,[51] mediante red teaming periodico e prove di robustezza integrate nella validazione pre-rilascio. La convergenza delle due tradizioni in un unico programma, ancorato al ciclo di esercitazioni del sistema di gestione della continuità ed eseguibile in ambienti di replica come i digital twin del capitolo 4, è il punto in cui sicurezza e continuità dei sistemi AI cessano di essere discipline separate. È anche, come mostra il capitolo seguente, la direzione verso cui converge il quadro regolatorio europeo.

[50] A. Basiri, N. Behnam, R. de Rooij, L. Hochstein, L. Kosewski, J. Reynolds, C. Rosenthal, «Chaos Engineering», IEEE Software, vol. 33, n. 3, 2016, pp. 35–41, DOI: 10.1109/MS.2016.60.

[51] National Institute of Standards and Technology, «Adversarial Machine Learning», NIST AI 100-2e2025, cit.

7. Governance e compliance

Quanto il capitolo precedente ha argomentato sul piano metodologico sta rapidamente diventando obbligo giuridico. Il quadro che ne risulta ha un'architettura a strati: una disciplina orizzontale del prodotto AI, l'AI Act; le discipline verticali della resilienza, NIS2 e CER per i soggetti essenziali e le entità critiche, DORA per il settore finanziario; e un livello di standard volontari che offre alle prime i veicoli attuativi. Questo capitolo ne esamina i contenuti rilevanti per la continuità operativa, che la Tabella 7.1 ricompone in matrice comparativa a chiusura del capitolo.

7.1 L'AI Act e i requisiti di robustezza per i sistemi high-risk

Per gli operatori di infrastrutture critiche il punto di ingresso nel regolamento è l'Allegato III, che classifica come ad alto rischio i sistemi di AI destinati a essere utilizzati come componenti di sicurezza nella gestione e nell'esercizio delle infrastrutture digitali critiche, del traffico stradale e della fornitura di acqua, gas, riscaldamento ed elettricità: una definizione che intercetta molti degli use case del capitolo 4. La classificazione attiva la catena dei requisiti degli articoli 9–15 (sistema di gestione del rischio, governance dei dati, record-keeping, sorveglianza umana, accuratezza, robustezza e cybersecurity), con applicabilità dal 2 agosto 2026 per i sistemi dell'Allegato III e sanzioni che, per la violazione di tali requisiti, possono raggiungere i 15 milioni di euro o il 3% del fatturato mondiale.[52]

Riletto dalla prospettiva della continuità, l'articolo 15 rivela tutta la sua portata. Il quarto paragrafo impone che i sistemi ad alto rischio siano quanto più possibile resilienti rispetto a errori, guasti e incoerenze, precisando che la robustezza può essere conseguita mediante soluzioni tecniche di ridondanza, incluse quelle che la norma denomina espressamente piani di backup o fail-safe: per la prima volta un requisito di continuità entra nel diritto europeo come proprietà di prodotto del sistema AI, e non soltanto come misura organizzativa dell'ente che lo utilizza. Il quinto paragrafo impone resilienza rispetto ai tentativi di alterazione da parte di terzi, richiedendo misure contro data poisoning, model poisoning, adversarial examples e attacchi alla riservatezza: l'adversarial testing delineato nel capitolo 6 diventa così presidio di conformità, non soltanto buona pratica. Completano il disegno la sorveglianza umana dell'articolo 14, che dà veste giuridica alla fascia di supervisione della Figura 4.1, il logging dell'articolo 12, che coincide con l'audit trail degli asset AI, e la segnalazione degli incidenti gravi dell'articolo 73, che impone ai processi di incident management di riconoscere e notificare anche l'incidente «di prodotto» accanto a quello di sicurezza.

7.2 NIS2 e CER: continuità per soggetti essenziali ed entità critiche

La direttiva NIS2 adotta un approccio all-hazards e colloca la continuità nel nucleo delle misure minime di gestione del rischio: l'articolo 21 richiede espressamente la continuità operativa,

[52] Regolamento (UE) 2024/1689 (c.d. AI Act), cit., in particolare Allegato III, punto 2, e artt. 9–15.

comprensiva di gestione dei backup, disaster recovery e gestione delle crisi, accanto alla sicurezza della supply chain e alle politiche di cifratura, mentre l'articolo 23 scandisce la notifica degli incidenti significativi in stadi successivi, dal preallarme entro ventiquattro ore alla notifica entro settantadue fino alla relazione finale, con la responsabilità delle misure posta in capo agli organi di gestione.[53] La rilevanza per il presente lavoro è immediata: i sistemi di artificial intelligence dell'ente rientrano a pieno titolo tra i sistemi informatici e di rete cui le misure si applicano, sicché gli asset censiti nel capitolo 6 ricadono già oggi nel perimetro dell'articolo 21, e i fornitori di servizi AI in quello della sicurezza della catena di approvvigionamento.

La direttiva CER guarda al medesimo obiettivo dall'angolo del servizio: le entità critiche devono valutare i rischi rilevanti di ogni natura, adottare misure e piani di resilienza proporzionati e notificare gli incidenti che perturbano in modo significativo la fornitura dei servizi essenziali entro ventiquattro ore, con relazione dettagliata nei trenta giorni successivi; l'attuazione italiana affida alle autorità settoriali competenti l'individuazione dei soggetti critici entro il 17 gennaio 2026.[54] Poiché la dimensione cibernetica è demandata alla NIS2, la CER cattura ciò che a questa sfugge: la dipendenza operativa del servizio da capability, incluse quelle di AI, la cui indisponibilità o degrado comprometterebbe la fornitura. Gli scenari costruiti nel capitolo 6, dall'indisponibilità simultanea del fornitore AI alla deriva silente dei modelli, trovano così collocazione naturale nella valutazione dei rischi che la direttiva impone.

7.3 DORA e la resilienza operativa digitale

Il regolamento DORA merita attenzione anche oltre il proprio perimetro settoriale, per due ragioni: molte infrastrutture finanziarie sono a loro volta entità critiche, e il regolamento costituisce il modello più prescrittivo di resilienza digitale oggi vigente, un laboratorio le cui soluzioni tendono a migrare negli altri settori. I suoi pilastri, applicabili dal 17 gennaio 2025, coprono il framework di gestione del rischio ICT, la classificazione e segnalazione armonizzata degli incidenti, il programma di test di resilienza operativa digitale spinto fino al threat-led penetration testing per i soggetti maggiori, la gestione del rischio da terze parti ICT con un regime europeo di oversight sui fornitori critici e la condivisione delle informazioni sulle minacce.[55] Nella prospettiva di questo lavoro DORA anticipa il trattamento dei servizi AI come servizi ICT di terzi, con registro contrattuale, strategie di uscita e attenzione esplicita al rischio di concentrazione che riecheggia l'analisi sistemica del capitolo 5; e il suo programma di test obbligatori offre il veicolo naturale in cui innestare gli scenari di indisponibilità e non affidabilità dell'AI costruiti nel capitolo 6.

[53] Direttiva (UE) 2022/2555 (c.d. direttiva NIS2), cit., in particolare art. 21 e art. 23.

[54] Direttiva (UE) 2022/2557 (c.d. direttiva CER), cit.; d.lgs. 4 settembre 2024, n. 134, cit.

[55] Regolamento (UE) 2022/2554 (c.d. regolamento DORA), cit.

7.4 Standard volontari: ISO/IEC 42001, ISO/IEC 23894, NIST AI RMF

Il livello volontario fornisce i veicoli attuativi. ISO/IEC 42001:2023 è il primo standard certificabile di sistema di gestione per l'artificial intelligence e adotta la struttura armonizzata delle norme ISO, il che ne consente l'integrazione con i sistemi già in esercizio ai sensi di ISO 22301 e ISO/IEC 27001: la continuità dei sistemi AI può così essere governata come estensione del sistema integrato esistente anziché come programma separato.[56] Lo affianca ISO/IEC 23894:2023, che declina sull'AI l'impianto di risk management della ISO 31000 e offre l'ancoraggio metodologico per le valutazioni del rischio che gli strumenti vincolanti richiedono.[57]

Sul versante statunitense, l'AI Risk Management Framework del NIST organizza la gestione del rischio nelle quattro funzioni Govern, Map, Measure e Manage e annovera tra le caratteristiche della trustworthy AI la coppia «secure and resilient», saldando esplicitamente sicurezza e resilienza nel medesimo requisito;[58] il profilo per la generative AI del luglio 2024 ne specializza le azioni su dodici categorie di rischio proprie dei sistemi generativi,[59] e nell'aprile 2026 il NIST ha avviato lo sviluppo di un profilo dedicato alla trustworthy AI nelle infrastrutture critiche, destinato a guidare gli operatori del settore nelle pratiche di gestione del rischio delle capability AI: un segnale della convergenza tematica in corso tra la disciplina dell'AI e quella della protezione delle infrastrutture.[60]

La lettura d'insieme della Tabella 7.1 mostra che gli strumenti, pur diversi per natura e perimetro, convergono sui medesimi presidi: inventario e classificazione degli asset, programmi di test, governo delle terze parti, notifica degli incidenti, sorveglianza umana, modalità operative degradate. Per gli operatori la conseguenza pratica è tangibile: l'organizzazione che attua l'impianto metodologico del capitolo 6 costruisce, con il medesimo investimento, la parte sostanziale della propria conformità ai quattro strumenti vincolanti. Il capitolo 8 porta a sintesi questa convergenza in un framework unitario.

Tabella 7.1 – Matrice comparativa dei framework normativi e volontari rispetto ai requisiti di continuità dei sistemi AI.

[56] International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 42001:2023 – Information technology – Artificial intelligence – Management system», dicembre 2023.

[57] International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 23894:2023 – Information technology – Artificial intelligence – Guidance on risk management», febbraio 2023.

[58] National Institute of Standards and Technology, «Artificial Intelligence Risk Management Framework (AI RMF 1.0)», NIST AI 100-1, gennaio 2023.

[59] National Institute of Standards and Technology, «Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile», NIST AI 600-1, luglio 2024, DOI: 10.6028/NIST.AI.600-1.

[60] National Institute of Standards and Technology, concept note per un «AI RMF Profile on Trustworthy AI in Critical Infrastructure», aprile 2026, disponibile su: https://www.nist.gov/itl/ai-risk-management-framework.

Strumento	Natura e ambito	Requisiti rilevanti per la continuità	Presa sull'AI
AI Act (Reg. UE 2024/1689)	Regolamento orizzontale vincolante; approccio per livelli di rischio	Robustezza con ridondanze e piani fail-safe (art. 15); logging (art. 12); sorveglianza umana (art. 14); segnalazione degli incidenti gravi (art. 73)	*Diretta: disciplina il sistema AI in quanto prodotto*
NIS2 (Dir. UE 2022/2555)	Direttiva; soggetti essenziali e importanti; d.lgs. 138/2024	Misure all-hazards incluse business continuity, gestione dei backup, disaster recovery e crisis management (art. 21); notifica a stadi 24/72 ore; sicurezza della supply chain	*Indiretta: l'AI tra i sistemi informatici e di rete dell'ente*
CER (Dir. UE 2022/2557)	Direttiva; entità critiche; d.lgs. 134/2024	Valutazione dei rischi e piani di resilienza; notifica degli incidenti rilevanti entro 24 ore; continuità del servizio essenziale	*Indiretta: le dipendenze AI del servizio entrano nel risk assessment*
DORA (Reg. UE 2022/2554)	Regolamento; settore finanziario; applicabile dal 17.1.2025	Framework di ICT risk management; test di resilienza fino al TLPT; oversight delle terze parti ICT critiche; classificazione e segnalazione degli incidenti	*Sui servizi AI come servizi ICT di terzi: contratti, exit strategy, concentrazione*
ISO/IEC 42001:2023	Standard volontario certificabile (AIMS)	Sistema di gestione del ciclo di vita AI, integrabile con ISO 22301 e ISO/IEC 27001 in un sistema integrato	*Diretta: governance organizzativa dell'AI*
ISO/IEC 23894:2023	Linea guida volontaria (impianto ISO 31000)	Metodo di identificazione, analisi e trattamento dei rischi AI lungo il ciclo di vita	*Diretta: risk assessment specifico per l'AI*
NIST AI RMF (AI 100-1; AI 600-1)	Framework volontario	Funzioni Govern, Map, Measure, Manage; caratteristica di trustworthiness «secure and resilient»; profilo per le infrastrutture critiche in sviluppo (2026)	*Diretta: pratiche e vocabolario di gestione del rischio AI*

8. Verso un framework di AI-resilient business continuity

I capitoli precedenti hanno prodotto i materiali; questo li ricompone in un framework operativo. Più che una disciplina inedita, si propone un livello di integrazione sopra i sistemi di gestione che le organizzazioni già possiedono o stanno adottando (continuità ai sensi di ISO 22301, sicurezza delle informazioni, il nascente sistema di gestione dell'AI): un modo unico di trattare l'artificial intelligence al servizio, insieme, della resilienza operativa e della conformità al quadro del capitolo 7.

8.1 Principi di progettazione

Dall'analisi condotta discendono sei principi. Il primo è la *simmetria dual-use*: ogni capability AI introdotta per rafforzare la resilienza entra automaticamente nel perimetro degli asset da proteggere: ciascun blocco dell'architettura della Figura 4.1 genera perciò una riga nel censimento della Tabella 6.1. Il secondo è il *degrado come stato di progetto*: le modalità operative con AI indisponibile o non fidata sono definite, dimensionate e documentate prima dell'evento, con trigger oggettivi e interruttori indipendenti dal sistema da disattivare. Il terzo è la *fiducia verificata nel continuo*: l'efficacia di un modello è una grandezza monitorata, non una proprietà acquisita, e la sua verifica combina metriche di deriva e prove avversarie.

Il quarto principio è l'*indipendenza del ripristino*: le capacità di recovery, dal rollback dei modelli alle comunicazioni di crisi, non devono dipendere dallo strumento la cui compromissione le ha rese necessarie, secondo la lezione del caso CrowdStrike. Il quinto è la *supervisione proporzionale all'autonomia*: l'intensità del controllo umano cresce con la reversibilità e l'impatto delle azioni delegate, in coerenza con l'articolo 14 dell'AI Act. Il sesto è la *conformità come sottoprodotto*: un unico impianto di controlli, progettato sui requisiti sostanziali, alimenta con la stessa evidenza documentale gli obblighi convergenti mappati nella Tabella 7.1, evitando la proliferazione di programmi paralleli.

8.2 Il modello operativo

Il framework organizza questi principi su tre livelli sovrapposti al ciclo di vita della continuità, come rappresentato nella Figura 8.1. Il livello superiore, *AI for resilience*, innesta nelle cinque fasi del ciclo le capability del capitolo 4: il supporto analitico alla BIA, la previsione a monte, la rilevazione e la risposta accelerate, l'esercitazione continua su repliche. Il livello inferiore, *resilience of AI*, applica alle stesse fasi i presidi del capitolo 6: il censimento degli asset con la criticità a due dimensioni, gli obiettivi estesi di ripristino con il model recovery time e il punto fidato, il monitoraggio della deriva con la supervisione umana, il rollback e il retraining con recovery indipendente, il programma convergente di chaos e adversarial testing con le esercitazioni a sistemi spenti. La cornice esterna è il livello di governance, che ancora l'insieme agli strumenti del capitolo 7 e ne trae i requisiti di logging, notifica e sorveglianza.

La dinamica del modello è il ciclo stesso: le lezioni delle esercitazioni e degli incidenti reali alimentano l'aggiornamento congiunto dei piani e dei modelli, chiudendo su cadenza continua il circuito di apprendimento che la pianificazione tradizionale confinava alla revisione annuale. Il valore pratico del framework sta nella sua funzione di lingua franca: offre a chi gestisce la continuità, a chi sviluppa i modelli, a chi presidia la sicurezza e a chi risponde della conformità una mappa unica su cui collocare responsabilità e controlli, là dove oggi ciascuna funzione opera con tassonomie proprie.

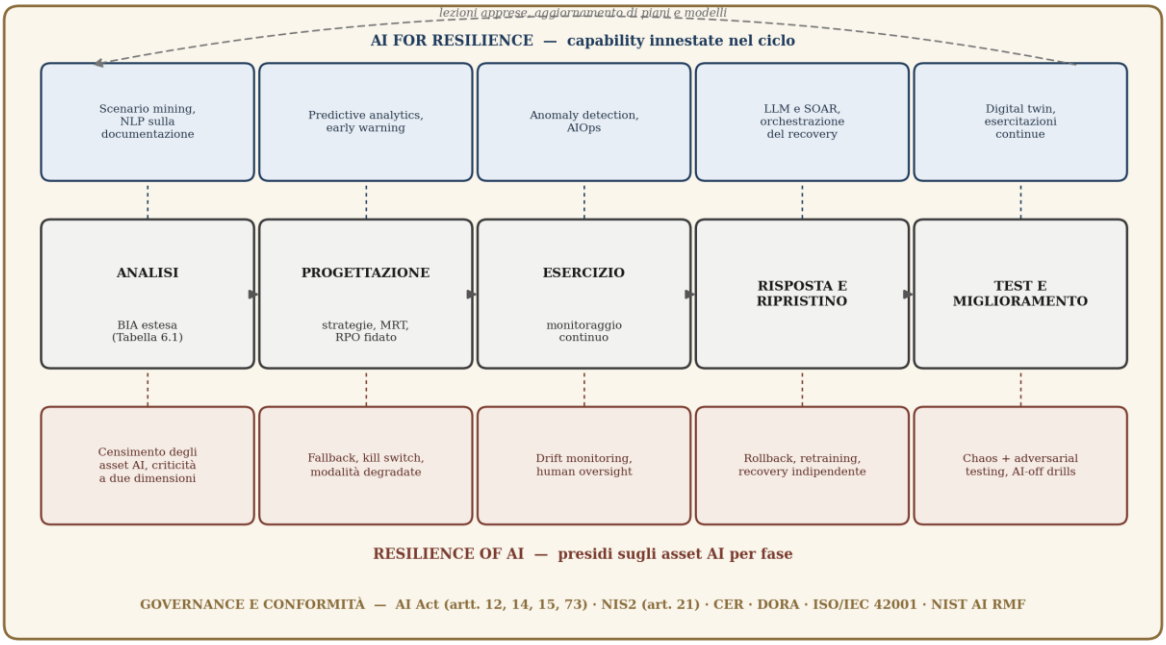

Figura 8.1 – Il framework di AI-resilient business continuity: capability abilitanti e presidi sugli asset AI integrati sulle fasi del ciclo di continuità, entro la cornice di governance e conformità.

8.3 Maturità e indicatori

Per orientare l'adozione è utile una scala di maturità a quattro livelli. Al primo livello, che si può dire *inconsapevole*, l'organizzazione impiega sistemi AI che i piani di continuità non censiscono. Al secondo, *censito*, gli asset AI figurano nella business impact analysis con obiettivi di ripristino definiti. Al terzo, *presidiato*, i fallback sono progettati e provati, la deriva è monitorata, il rollback è pronto e le terze parti AI hanno una exit strategy verificata. Al quarto, *integrato*, un unico programma di test copre continuità e sicurezza dei sistemi AI, le esercitazioni a sistemi spenti sono ricorrenti e l'evidenza prodotta alimenta direttamente la conformità.

La misurazione segue la stessa logica sostanziale: contano gli indicatori che descrivono capacità provate, non dichiarate. Tra i più significativi si segnalano la copertura del censimento, ossia la quota di sistemi AI in esercizio dotati di scheda di impatto; lo scarto tra model recovery time dichiarato e misurato nell'ultima prova; l'età dell'ultimo punto di ripristino la cui integrità sia stata effettivamente verificata; la frequenza e l'esito delle esercitazioni a sistemi spenti; la quota di modelli in produzione con rollback provato; il tempo di attivazione delle modalità

degradate osservato nei game day; la percentuale di fornitori AI critici con fallback collaudato. Sono grandezze semplici, ma la loro semplice richiesta costringe l'organizzazione a costruire le capacità che pretendono di misurare.

9. Sfide aperte e direzioni di ricerca

Tre nodi restano aperti e delimitano l'agenda di ricerca. Il primo riguarda la spiegabilità sotto pressione: le tecniche di explainability sono state sviluppate per contesti di analisi, mentre la gestione delle crisi richiede spiegazioni fruibili in minuti da operatori sotto stress, sufficienti a decidere se fidarsi di una raccomandazione automatica o attivare il downgrade. A questo si lega il problema della misurazione: lo stesso AI Act demanda alla Commissione, in cooperazione con gli organismi di normazione e di metrologia, lo sviluppo di benchmark e metodologie per accuratezza e robustezza,[61] ma per la resilienza dei sistemi AI in scenari di continuità non esistono ancora metriche condivise né dataset pubblici di riferimento, e la valutazione indipendente resta episodica rispetto alla letteratura di matrice industriale.

Il secondo nodo è l'AI agentica applicata al recovery. Agenti capaci di pianificare ed eseguire end-to-end sequenze di ripristino promettono di comprimere il mean time to recovery oltre i limiti dell'automazione a runbook, ma spostano sull'agente la stessa concentrazione di privilegio e di autonomia che questo lavoro ha indicato come pattern di rischio; la tassonomia NIST include già la sicurezza degli AI agent tra le aree della adversarial machine learning generativa,[62] ma mancano modelli consolidati per i confini di delega, per il contenimento delle azioni irreversibili e per interruttori di circuito che reggano anche alla compromissione dell'agente stesso. È il terreno su cui la promessa del capitolo 4 e il rischio del capitolo 5 si toccano più da vicino.

Il terzo nodo è umano e organizzativo. La conservazione delle competenze manuali sotto automazione prolungata, la calibrazione della fiducia degli operatori nei sistemi che li assistono e l'efficacia reale delle esercitazioni a sistemi spenti sono questioni empiriche su cui la letteratura offre principi ma pochi dati longitudinali. Gli obblighi di notifica introdotti dal quadro europeo, dall'articolo 73 dell'AI Act ai regimi NIS2 e CER, produrranno nei prossimi anni una base informativa senza precedenti sugli incidenti che coinvolgono sistemi AI: farne una fonte per la ricerca, nel rispetto della riservatezza, è forse la singola opportunità più concreta per dare fondamento empirico a una disciplina che ne ha urgente bisogno.

[61] Regolamento (UE) 2024/1689 (c.d. AI Act), cit., art. 15, par. 2.

[62] National Institute of Standards and Technology, «Adversarial Machine Learning», NIST AI 100-2e2025, cit.

10. Conclusioni

Il lavoro ha esaminato il rapporto tra artificial intelligence e continuità operativa delle infrastrutture critiche adottando una prospettiva dual-use, e le risposte alle domande di ricerca poste nel capitolo 1 si lasciano ora enunciare con nettezza. Alla prima: le capability abilitanti sono reali e documentate lungo l'intero ciclo della continuità, dalla manutenzione predittiva all'orchestrazione del ripristino, ma con maturità disomogenea, massima nella rilevazione e ancora prevalentemente assistiva nella decisione. Alla seconda: i rischi introdotti si organizzano in tre famiglie, le minacce esogene potenziate, le fragilità endogene dei modelli e il rischio sistemico di concentrazione, e il caso CrowdStrike mostra che l'amplificazione da automazione concentrata non è un'ipotesi ma un precedente.

Alla terza domanda il capitolo 6 ha risposto mostrando che gli strumenti consolidati reggono, a patto di estenderli: la business impact analysis acquisisce la seconda domanda sul degrado, gli obiettivi di ripristino si sdoppiano nel model recovery time e nel punto fidato, l'impianto MLOps si rivela l'infrastruttura di disaster recovery dei sistemi AI, le modalità degradate e il testing convergente completano il metodo. Alla quarta: il quadro europeo non solo consente ma impone questa estensione, con l'articolo 15 dell'AI Act che eleva backup e fail-safe a requisito di prodotto e con NIS2, CER e DORA che ne fanno obbligo organizzativo; gli standard volontari forniscono i veicoli, e la convergenza dei presidi rende la conformità un sottoprodotto del lavoro ben fatto anziché un programma parallelo.

Il messaggio per gli operatori è in fondo tradizionale, ed è questa la sua forza: l'artificial intelligence va trattata con la stessa disciplina riservata a ogni asset critico. Censirla, assegnarle obiettivi di ripristino, progettarne l'assenza, provarla regolarmente: l'era dell'AI non abroga i fondamentali della continuità operativa, ne estende il perimetro. Restano i limiti propri di una rassegna condotta su un campo in rapida evoluzione: il framework proposto è una sintesi ragionata che attende validazione sul campo, e le fonti più recenti sono destinate a invecchiare in fretta. È un invito, più che una riserva: alle organizzazioni, a sperimentare l'impianto e a misurarne gli esiti; alla ricerca, a colmare i vuoti empirici indicati; a entrambe, a farlo prima che sia il prossimo incidente a dettare l'agenda.

Bibliografia

Letteratura scientifica

Alcaraz C., Lopez J., «Digital Twin: A Comprehensive Survey of Security Threats», IEEE Communications Surveys & Tutorials, vol. 24, n. 3, 2022, pp. 1475–1503, DOI: 10.1109/COMST.2022.3171465.

Basiri A., Behnam N., de Rooij R., Hochstein L., Kosewski L., Reynolds J., Rosenthal C., «Chaos Engineering», IEEE Software, vol. 33, n. 3, 2016, pp. 35–41, DOI: 10.1109/MS.2016.60.

Bayram F., Ahmed B. S., Kassler A., «From concept drift to model degradation: An overview on performance-aware drift detectors», Knowledge-Based Systems, vol. 245, 108632, 2022, DOI: 10.1016/j.knosys.2022.108632.

Beaman C., Barkworth A., Akande T. D., Hakak S., Khan M. K., «Ransomware: Recent advances, analysis, challenges and future research directions», Computers & Security, vol. 111, 102490, 2021, DOI: 10.1016/j.cose.2021.102490.

Cheng Q., Sahoo D., Saha A., Yang W., Liu C., Woo G., Singh M., Savarese S., Hoi S. C. H., «AI for IT Operations (AIOps) on Cloud Platforms: Reviews, Opportunities and Challenges», arXiv:2304.04661, 2023.

Dennis C. R., Evans C. S., Duckworth K., Skinner M. M., Hanna J., Thompson T., Herring D., Medford R. J., «Resilience in the Face of Disruption: Viewpoint on the CrowdStrike Incident in July 2024», JMIR Medical Informatics, vol. 13, e69958, 2025, DOI: 10.2196/69958.

Habibzadeh A., Feyzi F., Ebrahimi Atani R., «Large Language Models for Security Operations Centers: A Comprehensive Survey», arXiv:2509.10858, 2025.

Hammar K., Stadler R., «Intrusion Tolerance for Networked Systems through Two-Level Feedback Control», in Proceedings of the 54th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2024), 2024, pp. 338–352.

Hammar K., Alpcan T., Lupu E. C., «Incident Response Planning Using a Lightweight Large Language Model with Reduced Hallucination», in Proceedings of the 33rd Annual Network and Distributed System Security Symposium (NDSS 2026), San Diego, 2026.

Huang L., Yu W., Ma W., Zhong W., Feng Z., Wang H., Chen Q., Peng W., Feng X., Qin B., Liu T., «A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions», ACM Transactions on Information Systems, vol. 43, n. 2, 2025, pp. 1–55, DOI: 10.1145/3703155.

Kreuzberger D., Kühl N., Hirschl S., «Machine Learning Operations (MLOps): Overview, Definition, and Architecture», IEEE Access, vol. 11, 2023, pp. 31866–31879, DOI: 10.1109/ACCESS.2023.3262138.

Landauer M., Onder S., Skopik F., Wurzenberger M., «Deep learning for anomaly detection in log data: A survey», Machine Learning with Applications, vol. 12, 100470, 2023, DOI: 10.1016/j.mlwa.2023.100470.

Lin X., Zhang J., Deng G., Liu T., Liu X., Yang C., Zhang T., Guo Q., Chen R., «IRCopilot: Automated Incident Response with Large Language Models», arXiv:2505.20945, 2025.

Mirsky Y., Lee W., «The Creation and Detection of Deepfakes: A Survey», ACM Computing Surveys, vol. 54, n. 1, 2021, pp. 1–41, DOI: 10.1145/3425780.

Parasuraman R., Manzey D. H., «Complacency and Bias in Human Use of Automation: An Attentional Integration», Human Factors, vol. 52, n. 3, 2010, pp. 381–410, DOI: 10.1177/0018720810376055.

Serradilla O., Zugasti E., Rodriguez J., Zurutuza U., «Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects», Applied Intelligence, vol. 52, n. 10, 2022, pp. 10934–10964, DOI: 10.1007/s10489-021-03004-y.

Fonti istituzionali e standard tecnici

European Union Agency for Cybersecurity (ENISA), «ENISA Threat Landscape 2025», ottobre 2025.

Financial Conduct Authority, «CrowdStrike outage: lessons for operational resilience», 31 ottobre 2024.

Financial Stability Board, «The Financial Stability Implications of Artificial Intelligence», 14 novembre 2024.

Financial Stability Board, «Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector», ottobre 2025.

International Organization for Standardization, «ISO 22301:2019 – Security and resilience – Business continuity management systems – Requirements», Ginevra, 2019.

International Organization for Standardization, «ISO 22313:2020 – Security and resilience – Business continuity management systems – Guidance on the use of ISO 22301», Ginevra, 2020.

International Organization for Standardization, «ISO 22300:2021 – Security and resilience – Vocabulary», Ginevra, 2021.

International Organization for Standardization, «ISO/TS 22317:2021 – Security and resilience – Business continuity management systems – Guidelines for business impact analysis», Ginevra, 2021.

International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 23894:2023 – Information technology – Artificial intelligence – Guidance on risk management», 2023.

International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 42001:2023 – Information technology – Artificial intelligence – Management system», 2023.

International Organization for Standardization / International Electrotechnical Commission, «ISO/IEC 27031:2025 – Cybersecurity – Information and communication technology readiness for business continuity», 2ª ed., 2025.

Microsoft, «Helping our customers through the CrowdStrike outage», The Official Microsoft Blog, 20 luglio 2024.

National Institute of Standards and Technology, «Contingency Planning Guide for Federal Information Systems», NIST Special Publication 800-34 Rev. 1, 2010.

National Institute of Standards and Technology, «Artificial Intelligence Risk Management Framework (AI RMF 1.0)», NIST AI 100-1, 2023.

National Institute of Standards and Technology, «Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile», NIST AI 600-1, 2024.

National Institute of Standards and Technology, «Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations», NIST AI 100-2e2025, 2025.

National Institute of Standards and Technology, concept note per un «AI RMF Profile on Trustworthy AI in Critical Infrastructure», aprile 2026.

U.S. Government Accountability Office, «Cyber Resiliency: CrowdStrike Outage Highlights Challenges», GAO-24-107733, settembre 2024.

Atti normativi

Direttiva (UE) 2022/2555 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativa a misure per un livello comune elevato di cibersicurezza nell'Unione (direttiva NIS2), GU L 333 del 27.12.2022.

Direttiva (UE) 2022/2557 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativa alla resilienza dei soggetti critici (direttiva CER), GU L 333 del 27.12.2022.

Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativo alla resilienza operativa digitale per il settore finanziario (regolamento DORA), GU L 333 del 27.12.2022.

Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act), GU L del 12.7.2024.

Decreto legislativo 4 settembre 2024, n. 134, attuazione della direttiva (UE) 2022/2557 relativa alla resilienza dei soggetti critici.

Decreto legislativo 4 settembre 2024, n. 138, recepimento della direttiva (UE) 2022/2555 (NIS2).