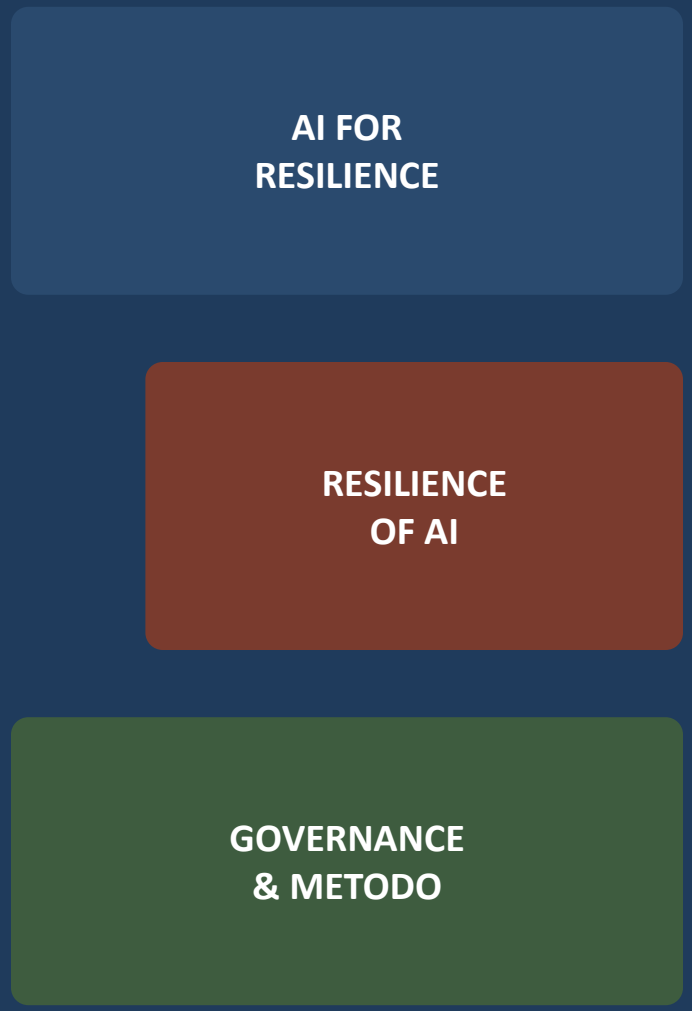
CORSO UNIVERSITARIO · MODULO DI CYBERSECURITY

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

Un approccio dual-use alla resilienza delle infrastrutture critiche

[Vincenzo Calabrò] · A.A. 2026/2027



INTRODUZIONE

Obiettivi di apprendimento

- Definire business continuity e disaster recovery e le metriche del BCM: BIA, MTPD, RTO, RPO, MBCO
- Mappare gli use case dell'AI sulle fasi del ciclo di continuità e sulle metriche che migliorano
- Classificare i rischi introdotti dall'AI: minacce esogene, fragilità endogene, rischi sistemici
- Applicare BIA e obiettivi di ripristino agli asset AI: modelli, dataset, pipeline, servizi di terzi
- Collocare i requisiti di AI Act, NIS2, CER e DORA nella pratica della continuità operativa

INTRODUZIONE

Il percorso della lezione

1. **Fondamenti**
   BC/DR, metriche, standard
2. **Il contesto**
   threat landscape e limiti dei piani tradizionali
3. **AI for resilience**
   le capability lungo il ciclo di continuità
4. **Resilience of AI**
   tre famiglie di rischio e il caso CrowdStrike
5. **Il metodo**
   l'AI come asset critico: BIA, MRT, MLOps, drills
6. **Governance**
   AI Act, NIS2, CER, DORA e il framework finale

19 LUGLIO 2024 · 04:09 UTC

Un aggiornamento, il mondo in blue screen

78 min

finestra di distribuzione dell’update difettoso

8,5 mln

dispositivi Windows in crash nel mondo

~30 h

di command center in un solo sistema sanitario

Nessuna AI in quell’aggiornamento. Ma il pattern (agente con privilegi profondi, flotta globale, aggiornamenti automatici fidati) è lo stesso con cui oggi distribuiamo le difese basate su AI e domani gli AI agent.

Il guasto viaggia a velocità software; il ripristino torna a velocità umana, dispositivo per dispositivo.

Fonti: GAO-24-107733; Microsoft (2024); Dennis et al., JMIR Medical Informatics (2025).

Business continuity e disaster recovery

Business continuity

- Capacità di continuare a erogare prodotti e servizi a livelli predefiniti durante un’interruzione (ISO 22301)
- Disciplina organizzativa: governata dal business continuity management (BCM) come sistema di gestione
- Copre persone, processi, fornitori e comunicazione di crisi, non solo la tecnologia

Disaster recovery

- Componente attuativa sul piano ICT: ripristino di sistemi informativi, dati e infrastrutture
- Strumenti tipici: siti alternativi, replica sincrona o asincrona, backup immutabili, failover orchestrato
- Inquadrato dalla NIST SP 800-34 in una famiglia coordinata di piani di contingency

In sintesi: il BCM decide che cosa deve sopravvivere e a quale livello; il disaster recovery costruisce il come sul piano tecnologico.

Le metriche: dalla BIA agli obiettivi di ripristino

BIA

Business impact analysis: identifica attività prioritarie, dipendenze e crescita degli impatti nel tempo

→

MTPD

Maximum tolerable period of disruption: oltre questa soglia gli impatti diventano inaccettabili

→

RTO · RPO

Tempo obiettivo di ripristino (RTO ≤ MTPD) e massima perdita di dati tollerabile

→

MBCO

Minimum business continuity objective: il livello minimo di servizio in modalità degradata

La logica temporale: la BIA determina l'MTPD; l'RTO deve stargli sotto; l'RPO governa frequenza dei backup e strategie di replica; l'MBCO definisce come si sopravvive nel frattempo. Tutto il resto della lezione è una rilettura di queste quattro grandezze.

Gli standard di riferimento

ISO 22301:2019

Requisiti del sistema di gestione, certificabile. Baricentro organizzativo e strategico; la tecnologia è una delle risorse da presidiare.

NIST SP 800-34

Guida tecnica alla contingency: sette passi, dalla policy al testing. Lessico operativo per dimensionare le soluzioni di ripristino.

ISO/IEC 27031:2025

ICT readiness for business continuity, 2ª edizione dopo 14 anni: condizioni di attacco attivo, cloud, terze parti; aggancio a MBCO, RTO, RPO.

Il punto cieco comune: nessuno tratta i sistemi AI come categoria autonoma di asset. Modelli, dataset e pipeline restano «generiche risorse ICT».

PARTE 2 · IL CONTESTO

Il threat landscape delle infrastrutture critiche

4.875

incidenti censiti da ENISA tra luglio 2024 e giugno 2025

53,7%

a carico dei soggetti essenziali definiti dalla NIS2

~60%

degli accessi iniziali passa dal phishing

OT

esposta su Internet: bersaglio ad alto valore per tutti gli attori

- Il ransomware resta la minaccia a maggiore impatto; gli attacchi DDoS di matrice hacktivista dominano per volume
- Le dipendenze digitali intrecciate propagano la perturbazione di un anello lungo l’intera supply chain dei servizi essenziali

Fonte: ENISA Threat Landscape 2025 (ottobre 2025).

Tre presupposti della pianificazione classica che cadono

Minacce lente

I piani si rivedono e si provano su cicli annuali; vulnerabilità e tecniche offensive evolvono in giorni.

Perimetri completi

Gli inventari di continuità non vedono gli asset algoritmici né i servizi cognitivi acquistati via API.

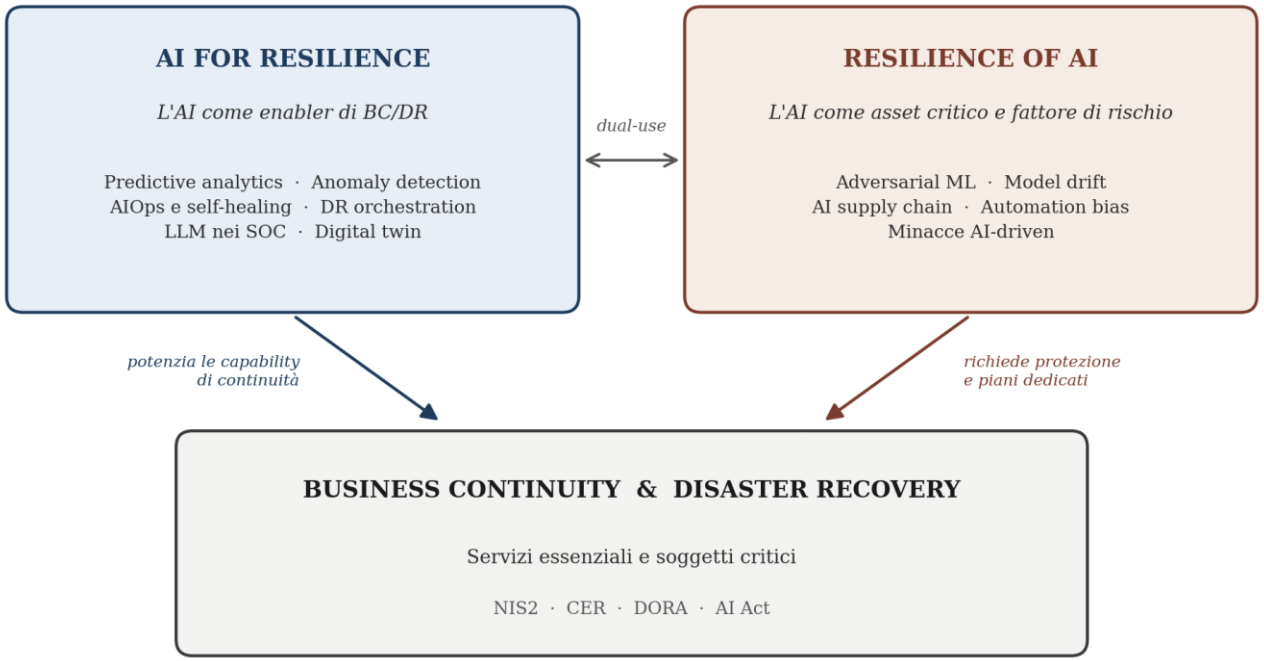
Reazione umana

Attacchi automatizzati e guasti a propagazione istantanea chiudono la finestra di intervento manuale.

La domanda della lezione: *l'AI può restituire ai difensori il vantaggio della velocità? E a quale prezzo in termini di nuove fragilità?*

LA TESI

La prospettiva dual-use



- AI for resilience: le tecniche di machine learning potenziano rilevazione, risposta e ripristino
- Resilience of AI: il sistema che protegge il business diventa esso stesso asset critico da proteggere
- Le due prospettive sono inseparabili: ogni capability introdotta apre superficie d'attacco e dipendenze

«Due facce di un'unica strategia di continuità»

La mappa: use case sul ciclo di continuità

Fase del ciclo	Use case AI	Metrica interessata
Analisi e pianificazione	NLP sulla documentazione, scenario mining	Copertura della BIA
Prevenzione	Predictive maintenance, previsione di saturazioni	Interruzioni evitate
Rilevazione	Anomaly detection su log, rete e OT; AIOps	MTTD
Risposta	LLM e SOAR: triage, sintesi, IR planning	MTTR
Ripristino	Orchestrazione del failover, self-healing	RTO e RPO effettivi
Esercitazione	Digital twin, stress test continui	Frequenza e realismo dei test

Tassonomia completa: Tabella 4.1 del white paper. Compiti AIOps: incident detection, failure prediction, root cause analysis, automated actions (Cheng et al., 2023).

Prevedere e rilevare: predictive analytics e AIOps

Predictive maintenance

- Modelli su vibrazioni, temperature e telemetrie stimano la remaining useful life dei componenti
- Effetto sulla continuità: il guasto subito diventa fermo programmato, con impatti negoziati
- Matura nei settori asset-intensive: energia, trasporti, ciclo idrico

Anomaly detection e AIOps

- Il deep learning sui log supera i metodi convenzionali su dati non strutturati e formati instabili
- Correlazione degli alert, failure prediction, root cause analysis: dal segnale grezzo alla diagnosi
- Beneficio primario: comprimere il MTTD e ridurre l'alert fatigue degli analisti

Contrappunto: qualità dei dati, deriva dei modelli e falsi positivi hanno costi di esercizio reali. Li riprendiamo nella parte 4.

Riferimenti: Serradilla et al., Applied Intelligence (2022); Landauer et al., Machine Learning with Applications (2023).

LLM nei SOC e nell’incident response

Lo sviluppo più recente e di più rapida diffusione

- Analisi dei log e triage degli alert accelerati; conoscenza specialistica on demand per analisti junior
- Sintesi degli incidenti e redazione assistita di situation report e comunicazioni di crisi
- Frontiera: incident response planning con tecniche di contenimento delle hallucination (NDSS 2026)
- Prototipi di risposta automatizzata end-to-end (IRCopilot): l’intero ciclo dell’incidente assistito

Condizione d’uso: hallucination e prompt injection restano aperte. L’impiego operativo corrente privilegia configurazioni assistive con validazione umana delle decisioni: l’LLM propone, l’analista dispone.

Riferimenti: Habibzadeh et al. (2025); Hammar, Alpcan, Lupu, NDSS 2026; Lin et al. (2025).

Chiudere il ciclo: dall’insight all’azione

Self-healing e intrusion tolerance

Runbook automatici: riavvio di servizi, scaling, isolamento di nodi. In ricerca: sistemi che mantengono il servizio sotto attacco con controllo a retroazione (DSN 2024).

DR orchestration

Failover orchestrato, coerenza continua tra siti, prove di recovery ricorrenti. Le azioni irreversibili restano sotto approvazione umana e shadow mode in adozione.

Backup difesi dal ransomware

Storage immutabile più AI: pattern di cifratura massiva (entropia, tassi di modifica), selezione dell’ultimo restore point integro, stima del perimetro compromesso.

Sui backup si difende l’RPO reale: l’ultimo punto disponibile non è sempre l’ultimo punto fidato.

Riferimenti: Hammar e Stadler, DSN 2024; Beaman et al., Computers & Security (2021); ISO/IEC 27031:2025.

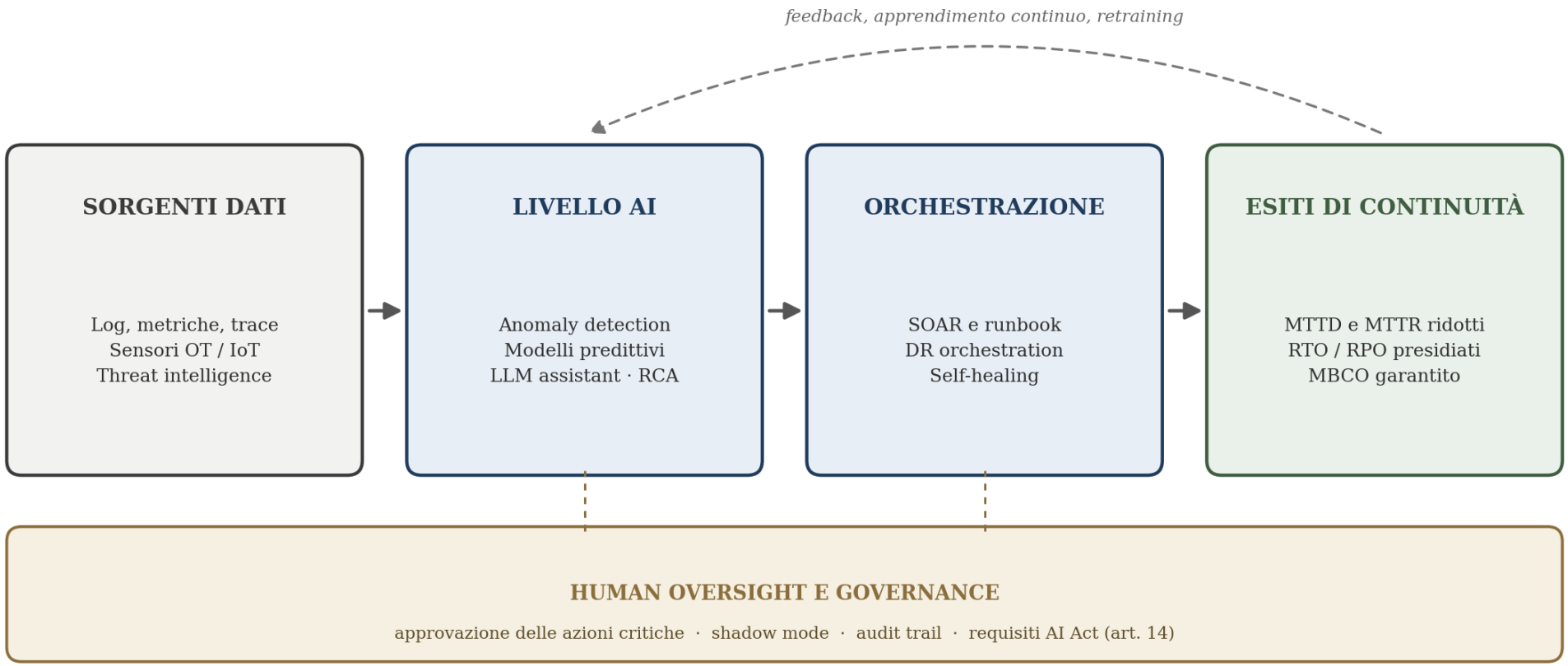
Digital twin: l’esercitazione diventa processo continuo

La replica virtuale connessa al sistema fisico

- Analisi what-if su guasti e attacchi senza toccare la produzione
- Stress test di capacità e validazione preventiva dei piani di ripristino
- Addestramento dei team di risposta su scenari realistici, anche generati con LLM (previa validazione esperta)
- Supera il limite del test annuale: rigiocare incidenti reali e varianti sintetiche, in continuo

Avvertenza dual-use: il twin è lo specchio fedele dell’impianto: va protetto con lo stesso rigore dell’originale, o diventa la mappa perfetta per l’attaccante (Alcaraz e Lopez, 2022).

L’architettura d’insieme



Il rovescio della medaglia: tre famiglie di rischio

Minacce esogene

L'AI nelle mani degli attaccanti: social engineering sintetico, deepfake, malware assistito

Fragilità endogene

Il comportamento dei modelli: drift, hallucination, automation bias degli operatori

Rischi sistemici

La struttura del mercato: concentrazione di fornitori, AI supply chain, common mode failure

Tratto comune: molte di queste minacce bersagliano le stesse capability difensive appena viste. Il sistema di protezione entra nel perimetro delle cose da proteggere.

Minacce AI-driven: l’inganno industrializzato

>80%

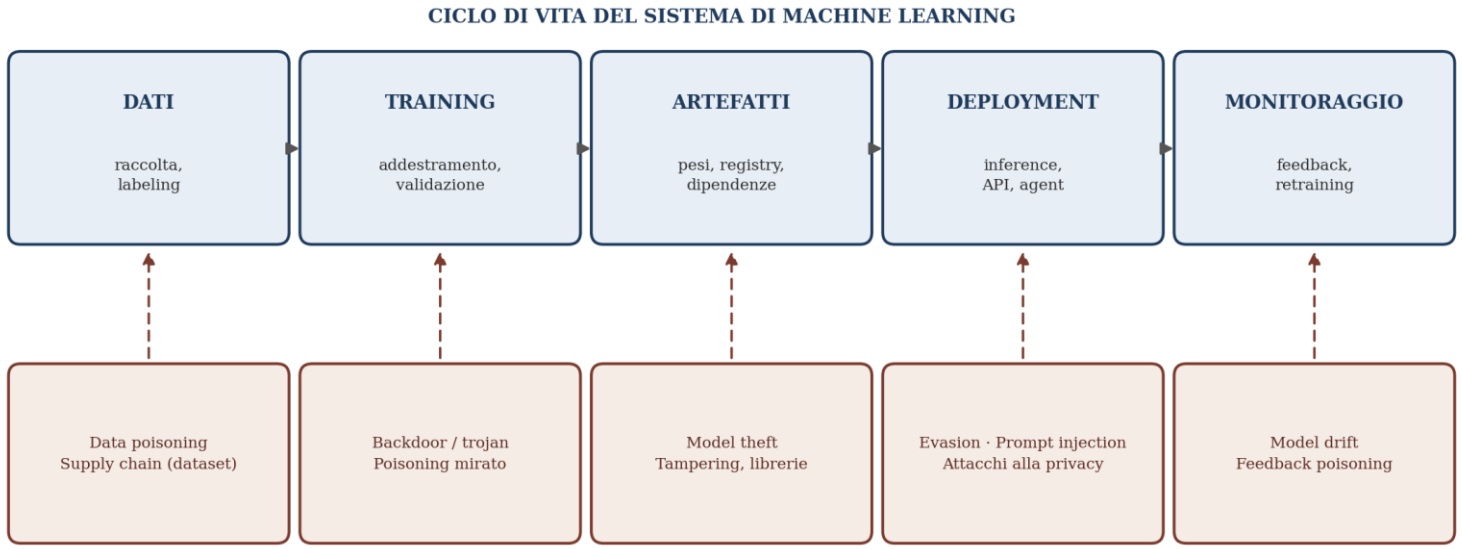
del social engineering osservato a inizio 2025
è phishing AI-supported

- LLM per malspam, spear phishing e vishing: qualità nativa, scala industriale, costo marginale nullo
- Deepfake audio-video in tempo reale: impersonation di dirigenti e fornitori nelle comunicazioni operative
- Malware assistito dai modelli generativi: cicli di sviluppo compressi, variazione continua contro le firme

Perché colpisce la continuità: i processi di crisi vivono di canali in deroga: telefonate di escalation, approvazioni urgenti, istruzioni di failover. È il terreno ideale dell’impersonation sintetica. Contromisura di piano: verifiche out-of-band e parole d’ordine procedurali che non transitino sul canale potenzialmente compromesso.

Fonti: ENISA Threat Landscape 2025; Mirsky e Lee, ACM Computing Surveys (2021).

Adversarial machine learning: la tassonomia NIST



Vettori di attacco per fase (classificazione: NIST AI 100-2e2025)

Il degrado silente

- Un sistema convenzionale compromesso si rompe; un modello avvelenato continua a rispondere, solo peggio
- Un assistente vittima di prompt injection esegue azioni legittime per fini ostili
- La detection si spegne gradualmente mentre i cruscotti restano verdi

Fonte: NIST AI 100-2e2025. PredAI: evasion, poisoning, privacy. GenAI: supply chain, prompt injection, sicurezza degli AI agent.

Fragilità intrinseche: nessun avversario richiesto

Model drift	I dati in esercizio si allontanano da quelli di addestramento: stagionalità, modifiche impiantistiche, nuovi pattern. Il modello invecchia in silenzio.	*Bayram et al., 2022*
Hallucination	Contenuti plausibili ma non fattuali: un'informazione falsa e verosimile entra nella catena decisionale nel momento di massima pressione.	*Huang et al., 2025*
Automation bias	L'affidabilità percepita induce compiacenza: i sistemi migliori generano la vigilanza peggiore verso l'errore raro che prima o poi arriva.	*Parasuraman e Manzey, 2010*

La combinazione critica: errori a velocità macchina incontrano una supervisione umana mal calibrata.

Rischio sistemico: concentrazione e common mode failure

- Pochi foundation model, poche piattaforme cloud, una filiera ristretta di hardware specializzato
- Accesso via API: indisponibilità, modifiche o compromissioni del fornitore si propagano istantaneamente a valle
- FSB (2024): dipendenze da terzi e concentrazione dei provider tra le vulnerabilità con potenziale sistemico
- Modelli simili commettono errori simili: il guasto individuale diventa common mode failure

Due scenari da portare in esercitazione: l'indisponibilità simultanea, multi-organizzazione, di un servizio AI esterno; la deriva coordinata di comportamenti automatizzati alimentati dagli stessi modelli.

Fonti: FSB, The Financial Stability Implications of AI (2024) e Monitoring Adoption of AI (2025); ENISA Threat Landscape 2025 sulla AI supply chain.

CASO DI STUDIO

Dentro l’outage CrowdStrike

I numeri sul campo (ECU Health)

- 04:09 UTC rilascio dell’update, 05:27 stop: 78 minuti di esposizione globale
- 8.647 blue screen, circa il 60% del parco dispositivi del sistema sanitario
- 729 macchine ripristinate a mano: safe mode, chiave BitLocker, rimozione del file
- Command center operativo per circa 30 ore; EHR salvato dalla finestra notturna di basso carico

Le lezioni regolatorie (FCA)

Identificare i single point of failure; testare gli aggiornamenti con rilasci progressivi; includere le terze parti negli scenari di resilienza.

La lettura prospettica

Agente privilegiato + flotta globale + update automatici fidati: è il pattern di deployment delle difese AI. Un model update difettoso o avvelenato avrebbe la stessa cinetica. Quindi: rollback indipendente dallo strumento guasto, canary anche per i modelli.

Fonti: GAO-24-107733; Dennis et al., JMIR Medical Informatics (2025); FCA (2024); Microsoft (2024).

L’AI come asset critico: la BIA a doppia domanda

Domanda classica: *che cosa si ferma se l’asset è indisponibile?*

Domanda nuova: *quali decisioni peggiorano se l’asset degrada senza fermarsi?*

Asset AI	Che cosa chiedersi in BIA
Modello (pesi, artefatto)	Autonomia delle decisioni? Impatto di un degrado non rilevato? Rollback vs retraining?
Dataset di training	Ricostruibili? Snapshot immutabili e lineage? Da quale data potrebbero essere avvelenati?
Feature store e pipeline	Quali sorgenti alimentano le feature? Che accade se una degrada in silenzio?
Registry, config, prompt	Chi può modificarli? Sono versionati, firmati, verificabili?
Servizio AI di terzi (API)	RTO contrattuale del fornitore? Esiste un fallback provato?

Il ripristino ripensato per i sistemi AI

Obiettivi estesi

- Model recovery time: dimensionato sullo scenario del retraining, non sull'ottimismo del rollback
- RPO = distanza dall'ultimo punto fidato, non dall'ultimo disponibile (il poisoning retrodata il punto utile)
- MLOps riletto: registry = backup del modello; pipeline riproducibile = procedura di ripristino; lineage = precondizione del punto integro

Esercizio e verifica

- Modalità degradate progettate prima, con trigger oggettivi (drift, confidenza) e MBCO definito
- Kill switch documentato e provato, indipendente dal sistema da spegnere
- AI-off drills: esercitazioni a sistemi spenti contro l'atrofia della competenza manuale
- Chaos engineering + adversarial testing in un unico programma di verifica

Riferimenti: Kreuzberger et al., IEEE Access (2023); Basiri et al., IEEE Software (2016); NIST AI 100-2e2025.

La stretta normativa europea

AI Act, art. 15: robustezza dei sistemi high-risk anche mediante ridondanze tecniche, inclusi «backup or fail-safe plans». Per la prima volta la continuità è requisito di prodotto del sistema AI. Il par. 5 impone misure contro data e model poisoning, adversarial examples, prompt injection: l'adversarial testing diventa conformità. Obblighi per l'Allegato III dal 2 agosto 2026.

NIS2

Continuità, backup, disaster recovery e crisis management tra le misure minime (art. 21); notifiche 24/72 h. In Italia: d.lgs. 138/2024.

CER

Piani di resilienza dei soggetti critici, notifica incidenti in 24 h. D.lgs. 134/2024; individuazione dei soggetti dal gennaio 2026.

DORA

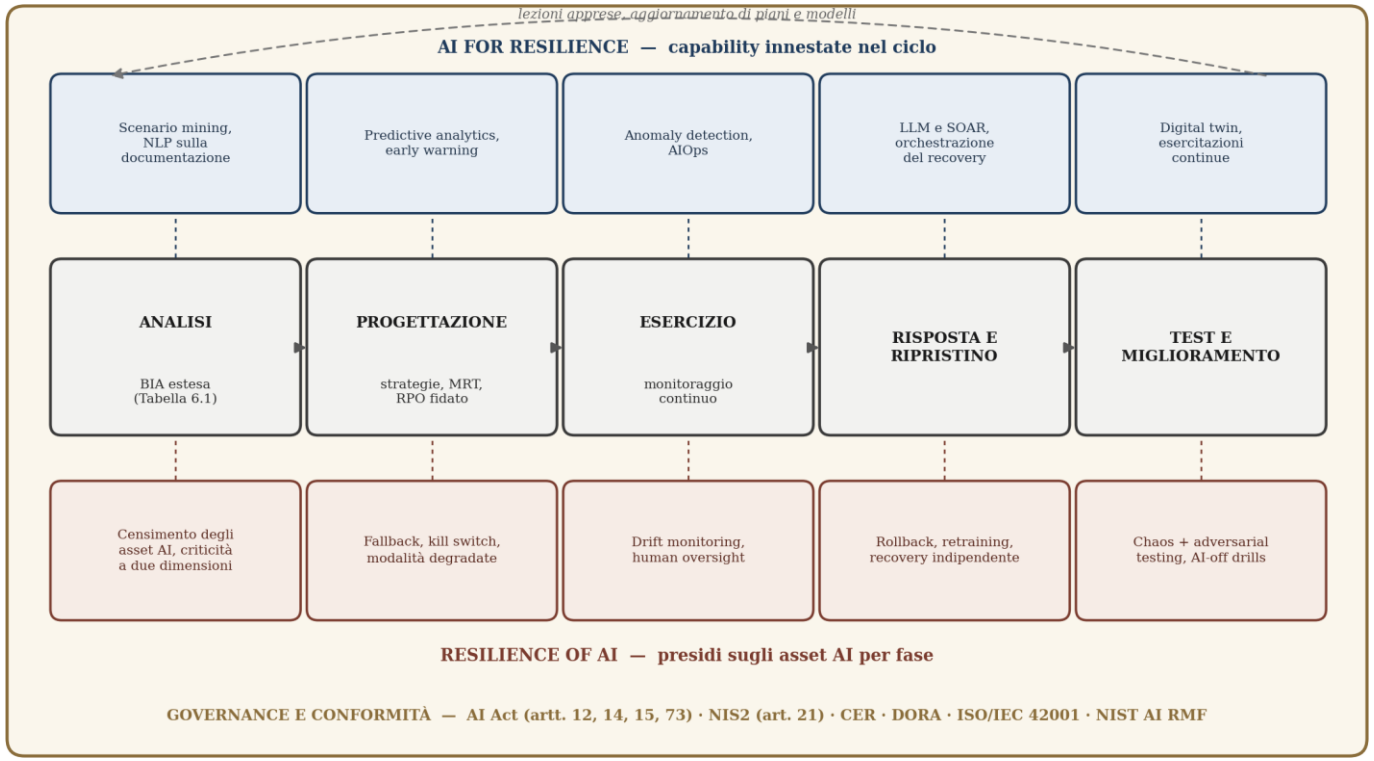
Settore finanziario, dal 17.1.2025: test fino al TLPT, oversight delle terze parti ICT critiche, exit strategy. Il laboratorio regolatorio.

Standard

ISO/IEC 42001 e 23894, NIST AI RMF: i veicoli attuativi. In sviluppo (2026) il profilo NIST per le infrastrutture critiche.

Gli strumenti convergono sugli stessi presidi: chi attua il metodo della parte 5 costruisce, con lo stesso investimento, gran parte della conformità.

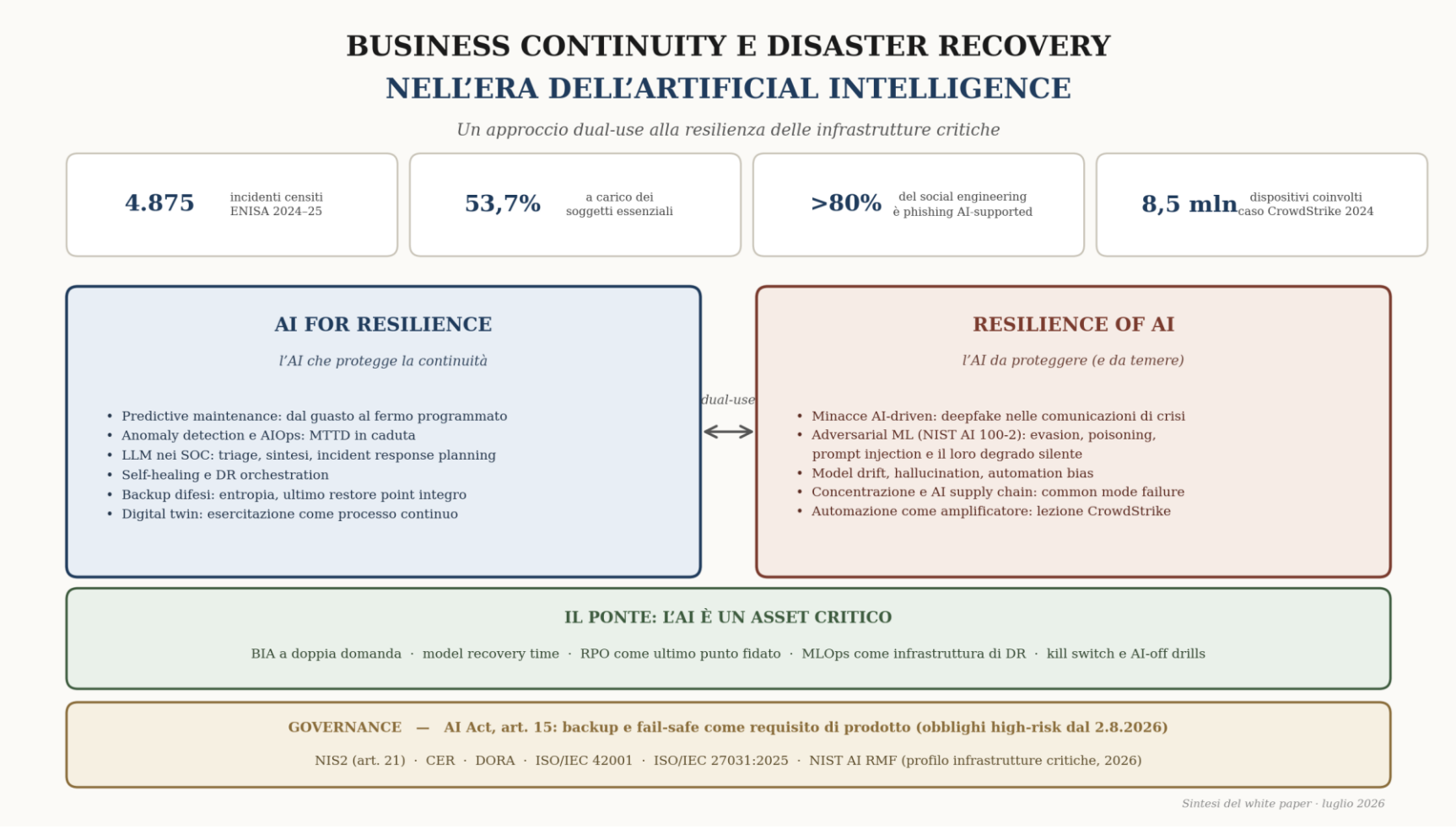
Il framework di AI-resilient business continuity



Maturità: inconsapevole → censito → presidiato → integrato. **Indicatori:** capacità provate, non dichiarate (MRT misurato, ultimo restore point verificato, fallback dei fornitori collaudato).

SINTESI

Il quadro in una pagina



IN AULA

Discussione ed esercitazione

Tre domande per la discussione

- Quale RTO ha senso per un modello che richiede retraining? Da quali fattori dipende davvero?
- Il vostro SOC adotta un assistente LLM: quali trigger di downgrade e quali validazioni umane definireste?
- Rigiochiamo CrowdStrike come model update avvelenato: che cosa cambia nella risposta e nel ripristino?

Esercitazione (gruppi, 45’): condurre una BIA su un sistema AI reale o simulato: censimento degli asset (modello, dati, pipeline, terze parti), doppia domanda, MRT e punto fidato, modalità degradata con trigger. Chiusura in plenaria con un tabletop «AI unavailable» di 15 minuti.

MESSAGGI CHIAVE

1. Dual-use inseparabile: ogni capability AI introdotta per la resilienza entra nel perimetro da proteggere
2. Il degrado silente è una nuova categoria di guasto: i controlli si spengono mentre i cruscotti restano verdi
3. Il ripristino non deve mai dipendere dallo strumento guasto: rollback indipendente, canary anche per i modelli
4. La conformità (AI Act, NIS2, CER, DORA) è il sottoprodotto del metodo, non un programma parallelo

«L'era dell'AI non abroga i fondamentali della continuità operativa: ne estende il perimetro.»

Censire l'AI · assegnarle obiettivi di ripristino · progettarne l'assenza · provarla regolarmente

PER APPROFONDIRE

Riferimenti essenziali

- ENISA, Threat Landscape 2025 (ottobre 2025)
- NIST, AI 100-2e2025: Adversarial Machine Learning (2025)
- NIST, AI RMF 1.0 (2023) e profilo GenAI AI 600-1 (2024)
- GAO-24-107733, Cyber Resiliency: CrowdStrike Outage (2024)
- FCA, CrowdStrike outage: lessons for operational resilience (2024)
- Dennis et al., JMIR Medical Informatics, 13:e69958 (2025)
- FSB, The Financial Stability Implications of AI (2024)
- Reg. (UE) 2024/1689 (AI Act), in particolare artt. 12, 14, 15, 73
- Dir. (UE) 2022/2555 (NIS2) e Dir. (UE) 2022/2557 (CER)
- Reg. (UE) 2022/2554 (DORA)
- ISO 22301:2019; ISO/IEC 27031:2025; ISO/IEC 42001:2023
- Cheng et al., AIOps on Cloud Platforms, arXiv (2023)
- Huang et al., ACM Trans. on Information Systems 43(2) (2025)
- Kreuzberger et al., IEEE Access 11 (2023)

Bibliografia completa (40 riferimenti, tutti verificati su fonte primaria) nel white paper allegato al corso.