Silvano Bari

AIIC – ISACA Roma
6 maggio 2025

WEBINAR
Critical Infrastructure Resilience and Artificial Intelligence

PROGRAMMA

15:00 - I GdL nella storia di AIIC *Silvano Bari*

15:10 – Applicazioni della Intelligenza Artificiale per la Resilienza delle Infrastrutture Critiche: sono solo vantaggi? *Sandro Bologna*

15:30 – Applicazioni della Intelligenza Artificiale tra passato, presente e futuro: cosa ci aspettiamo? *Alberto Stefanini*

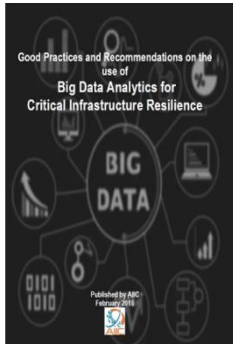
15:50 – Vincoli Etici e Sociali nell'utilizzo della Intelligenza Artificiale nelle Infrastrutture Critiche: sono soltanto degli impedimenti? *Luigi Carrozzi*

16:10 – Come affrontare il tema della valutazione del rischio nelle Infrastrutture Critiche che fanno uso della Intelligenza Artificiale: possiamo farne a meno? *Francesca Della Mea*

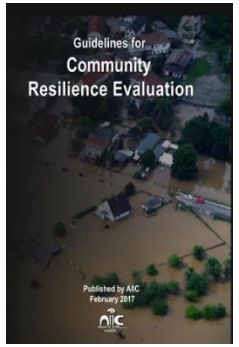
16:30 – Vincoli Normativi ed Etici non bastano ad assicurare la Resilienza: il caso Heathrow e un auspicabile contributo della intelligenza artificiale *Alberto Caruso De Carolis*

16:50 – Conclusioni: idee per un nuovo mondo *Luisa Franchina*

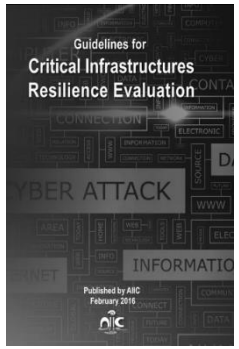
17:00 – Chiusura del Webinar



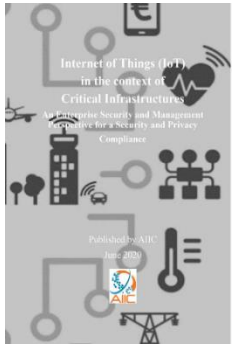
Good Practices and
Recommendations
on the Use of
Big Data Analytics for
Critical Infrastructure Resilience



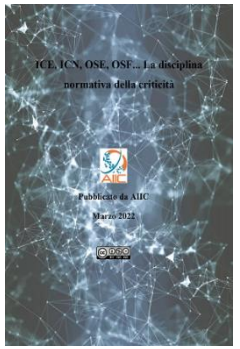
Guidelines for
Community
Resilience
Evaluation



Guidelines for
Critical Infrastructure
Resilience Evaluation



Internet of Things
In the Context of
Critical Infrastructure

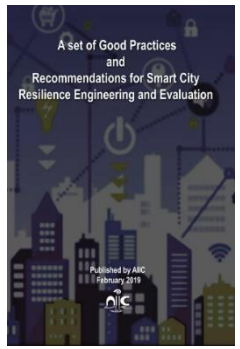


ICE, ICN, OSE, OSF…
La disciplina normativa
della Criticità

AIIC Gruppi di Studio e Pubblicazioni

www.infrastrutturecritiche.it

La Direttiva (UE)
2022/2555 NIS 2
e la Cybersecurity

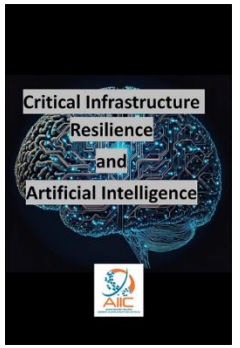


A set of Good Practices and
Recommendations
for Smart City
Resilience Engineering
and Evaluation

La Direttiva (UE)
2022/2557:
La Resilienza dei
Soggetti Critici
Peculiarità e parallelismi
con alcuni settori dei trasporti:
Aviation e Maritime

La Protezione
degli Spazi Pubblici
Analisi degli aspetti organizzativi,
delle tecnologie per la mitigazione
del rischio e dei loro vincoli

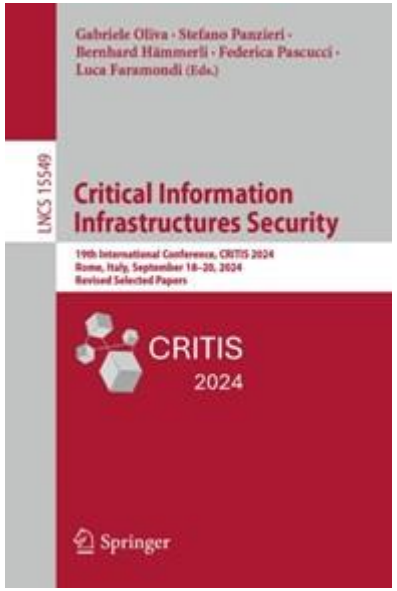
Resilienza delle
Infrastrutture Critiche
e Cambiamenti
Climatici



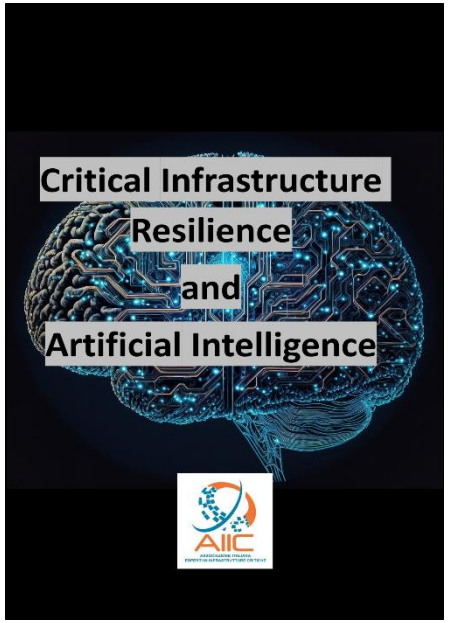
Critical Infrastructure
Resilience
and
Artificial Intelligence

AIIC Gruppi di Studio e Pubblicazioni

Quaderni IRCRES
del CNR
Cambiamento Climatico e Sostenibilità:
una visione multidisciplinare



Atti della 19^
Conferenza Internazionale
CRITIS 2024
Critical Information
Infrastructures Security



The Report examines the role and impact of Artificial Intelligence (AI) in critical infrastructure, providing a global overview of strategies, standards, ethical considerations, and regulatory frameworks. The aim is to identify best practices and offer recommendations for promoting resilience and sustainability through the responsible use of AI.
The document examines current policies and regulations governing AI use.
It identifies regulatory gaps and proposes improvements to balance innovation and safety.

Silvano Bari

Glauco Bertocchi

Sandro Bologna
coordinatore

Luigi Carrozzi

Francesca Della Mea

Adriana Peduto

Members of GdL AIIC

Critical Infrastructure Resilience and Artificial Intelligence

Maria Beatrice Versaci

Luisa Franchina

Alberto Stefanini

Raffaella D'Alessandro

Giorgio Pizzi

Alberto Caruso De Carolis

Sandro Bologna

AIIC – ISACA Rome Chapter

06 Maggio 2025

Critical Infrastructure Resilience and Artificial Intelligence

Silvano Bari, Glauco Bertocchi, Sandro Bologna*, Luigi Carrozzi, Alberto Caruso DeCarolis, Raffaella D'Alessandro, Francesca Della Mea, Luisa Franchina, Adriana Peduto, Giorgio Pizzi, Alberto Stefanini, Maria Cristina Versaci

AIIC Associazione Italiana esperti in Infrastrutture Critiche, Rome, Italy

**Corresponding author*

s.bologna@infrastrutturecritiche.it

From where the Report starts

The Report adopts guiding from different documents, in particular:

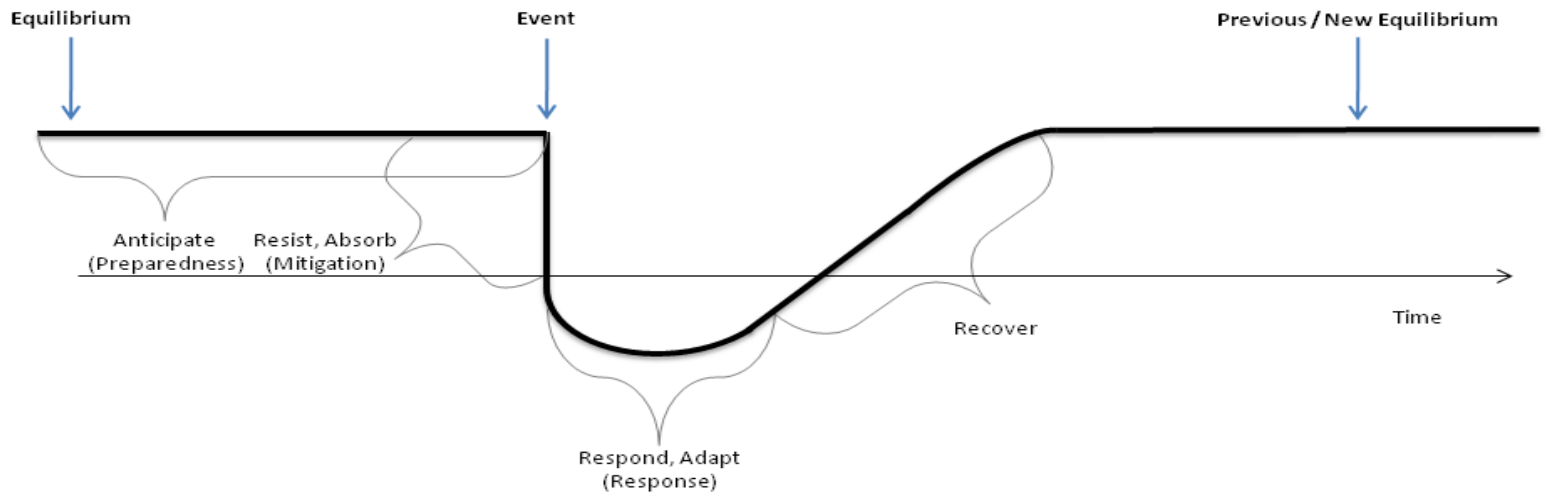
1. *Definition and classification of Critical Infrastructure Resilience as defined in the Directive (EU) 2022/2557 of the European Parliament and of the Council of 14 December 2022 on the resilience of critical entities.*
2. *European Commission through its High Artificial Intelligence Level Group (AI HLEG) has provided a valuable guiding document "Ethics Guidelines For Trustworthy AI" providing direction to adopt ethic principles in AI systems.*

It doesn't recommend any specific AI platform because of the continuous evolving of the AI technology

From where the Report starts

Graphycal Representation of Resilience

- A negative event occurres (i.e. a threat exploits the vulnerability of a given component)
- The system detects the occurrency of negative event
- The system reacts to the negative event and tries to recover its state

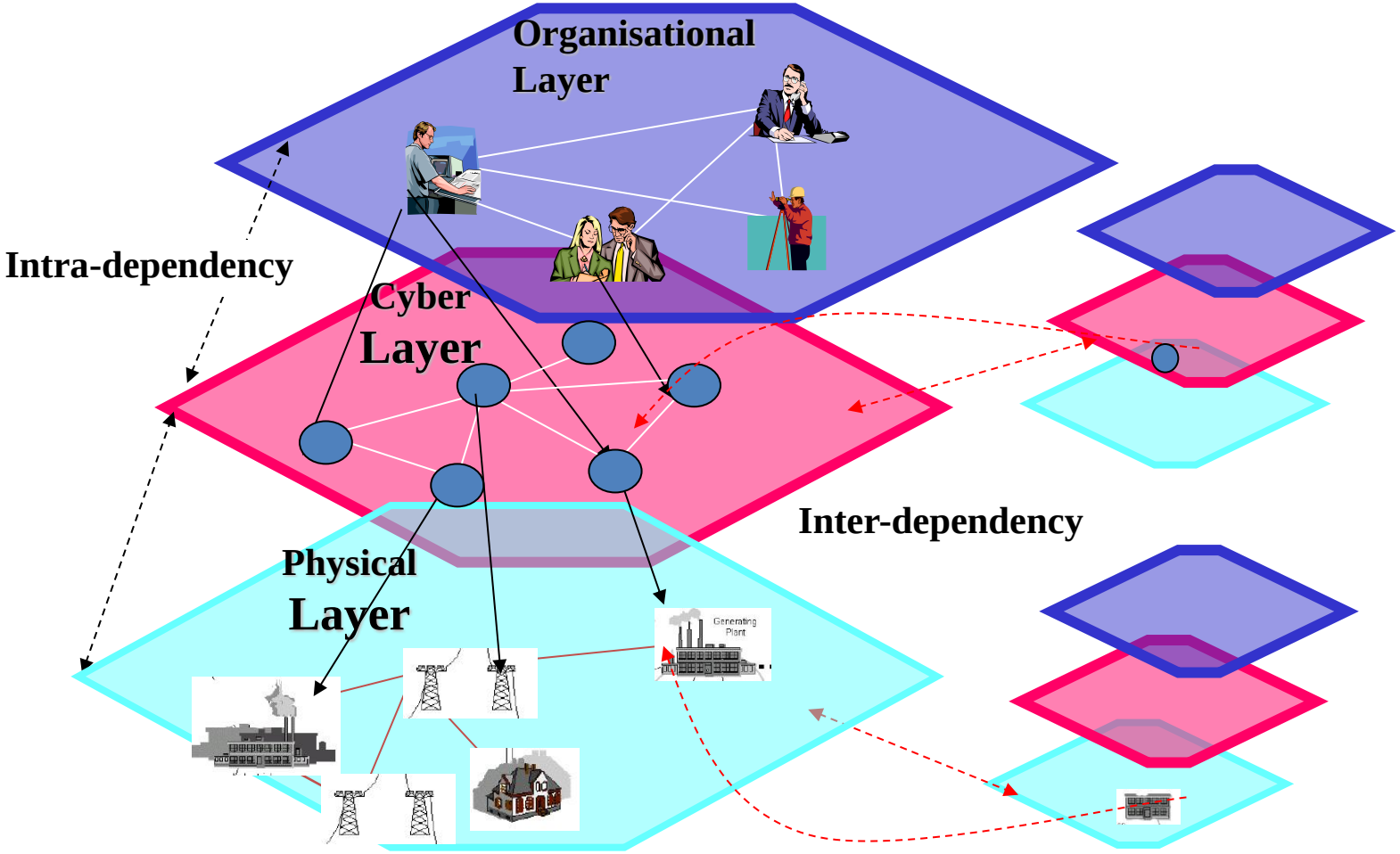


Source: ANL/DIS-12-1 Resilience: Theory and Applications

Guidelines for Critical Infrastructure Resilience Evaluation – AIIC 2016

From where the Report starts

Three Layers Model for the Infrastructure



Guidelines for Critical Infrastructure Resilience Evaluation – AIIC 2016

From where the Report starts

A Critical Infrastructure is not only made of technologies but especially of people, processes and organizations.
Resilience must take in consideration all these components, plus cultural background, to be complete and successful.

Source: Adapted from the USC Marshall School of Business
Institute for Critical Information Infrastructure Protection

Guidelines for Critical Infrastructure Resilience Evaluation – AIIC 2016

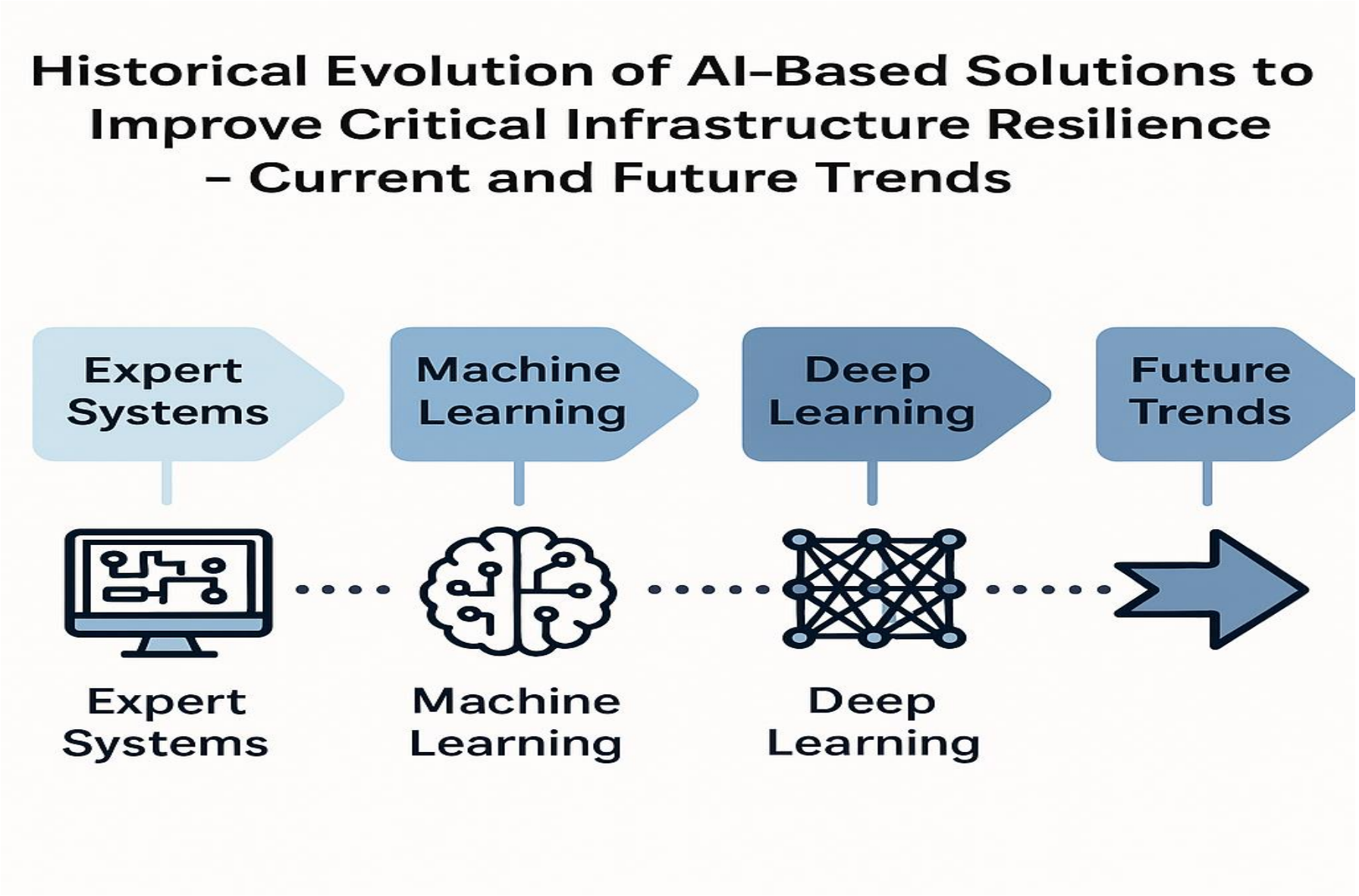
AI for what?

Predictive Maintenance: AI can analyze vast amounts of data from sensors embedded in critical infrastructure systems to predict potential failures before they occur.

Risk Management: AI algorithms can analyze historical data and current conditions to assess the potential risks to critical infrastructure from natural disasters, cyber attacks, or other threats.

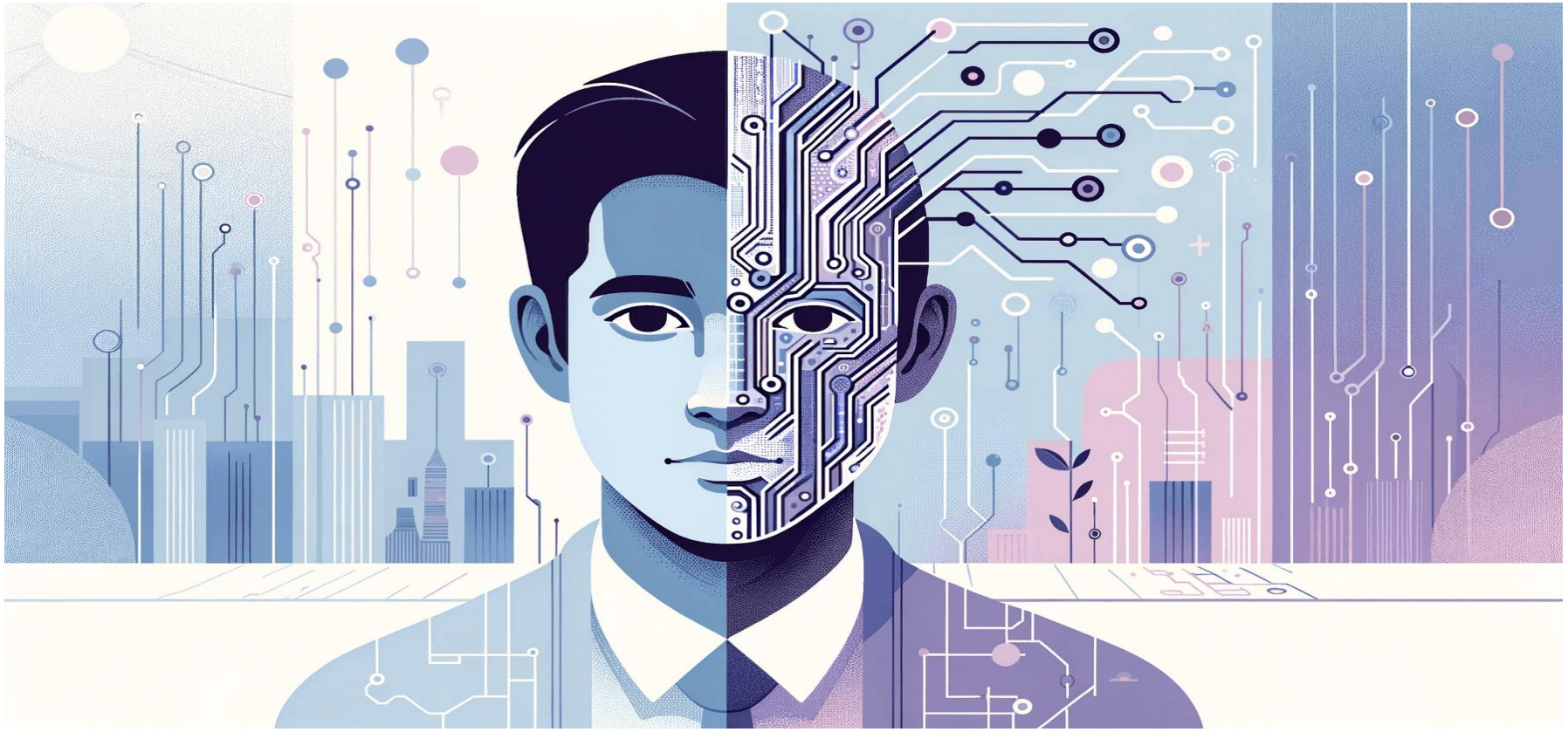
Cybersecurity: AI can enhance Security and Safety and is widely used for cybersecurity of critical infrastructure by detecting and responding to cyber threats in real-time.

Historical Evolution of AI in Critical Infrastructure



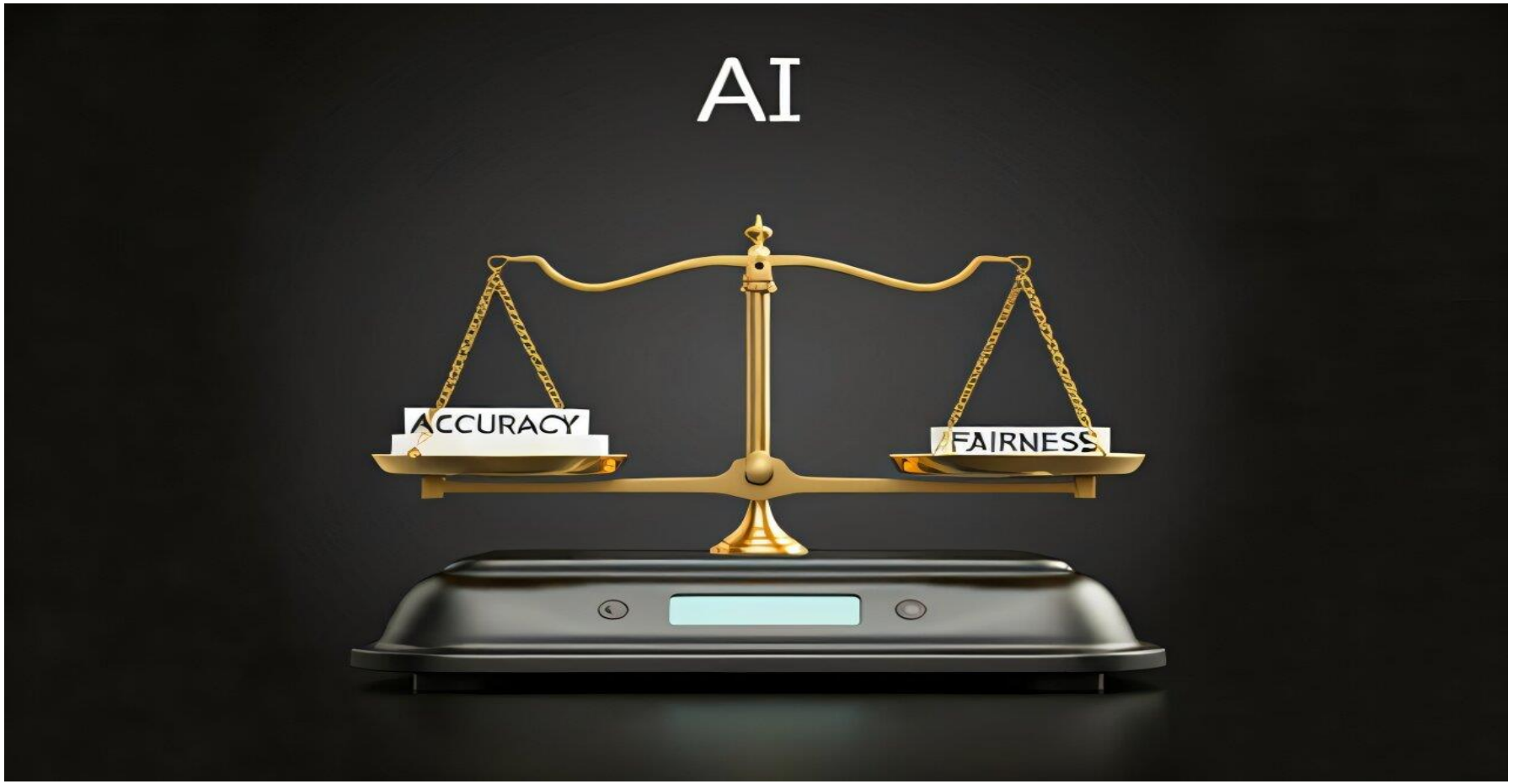
Governance and Regulatory Models
Transparency, Fairness, and Bias in AI Models

Transparency



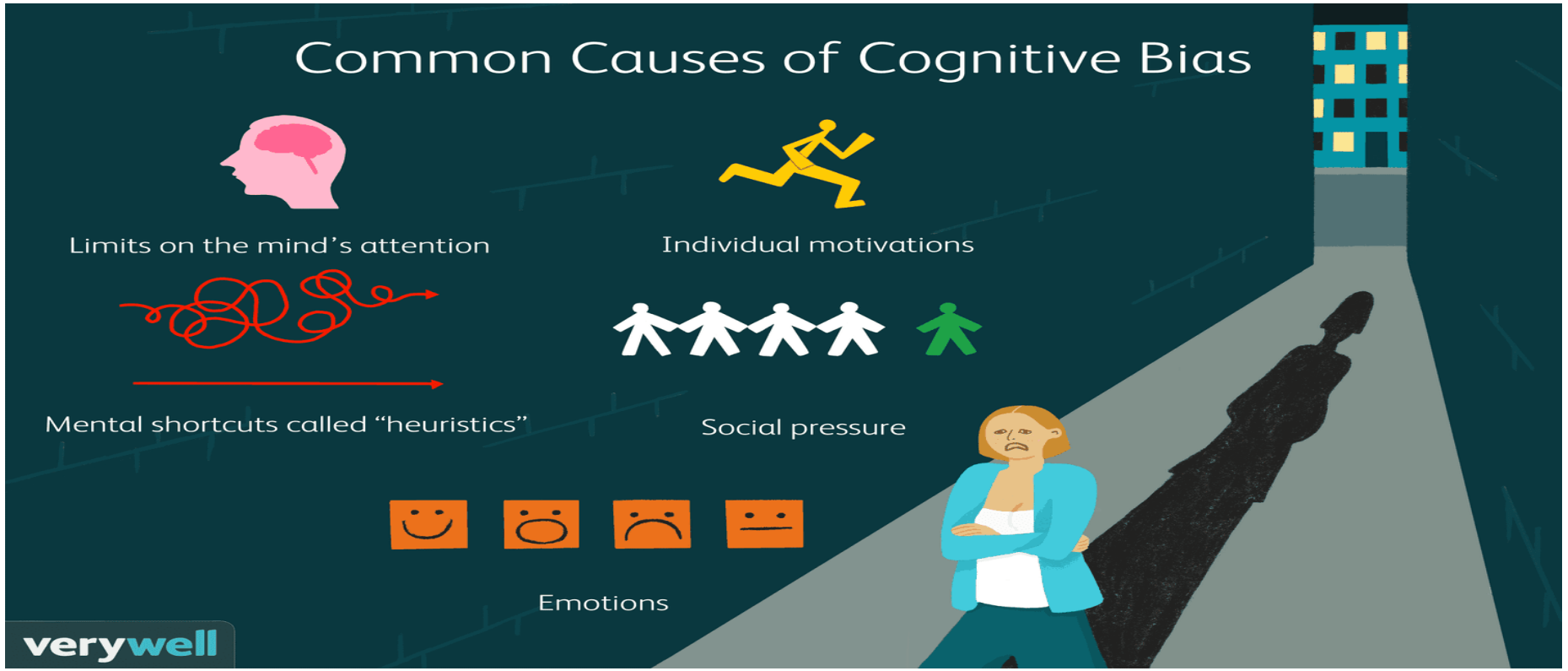
The concept of **transparency in AI models** refers to making the processes, decision-making, and outcomes of AI systems clear and understandable to users, developers, and stakeholders.

Fairness



Fairness in AI refers to the principle that artificial intelligence systems should make decisions **without bias**, treating all individuals and groups **equitably**.

Cognitive Bias



Cognitive bias in AI refers to **systematic errors** in machine learning models that arise from biased data, flawed algorithms, or human influence. These biases can lead to **unfair, inaccurate, or misleading** AI decisions, affecting various domains such as hiring, healthcare, finance, and law enforcement.

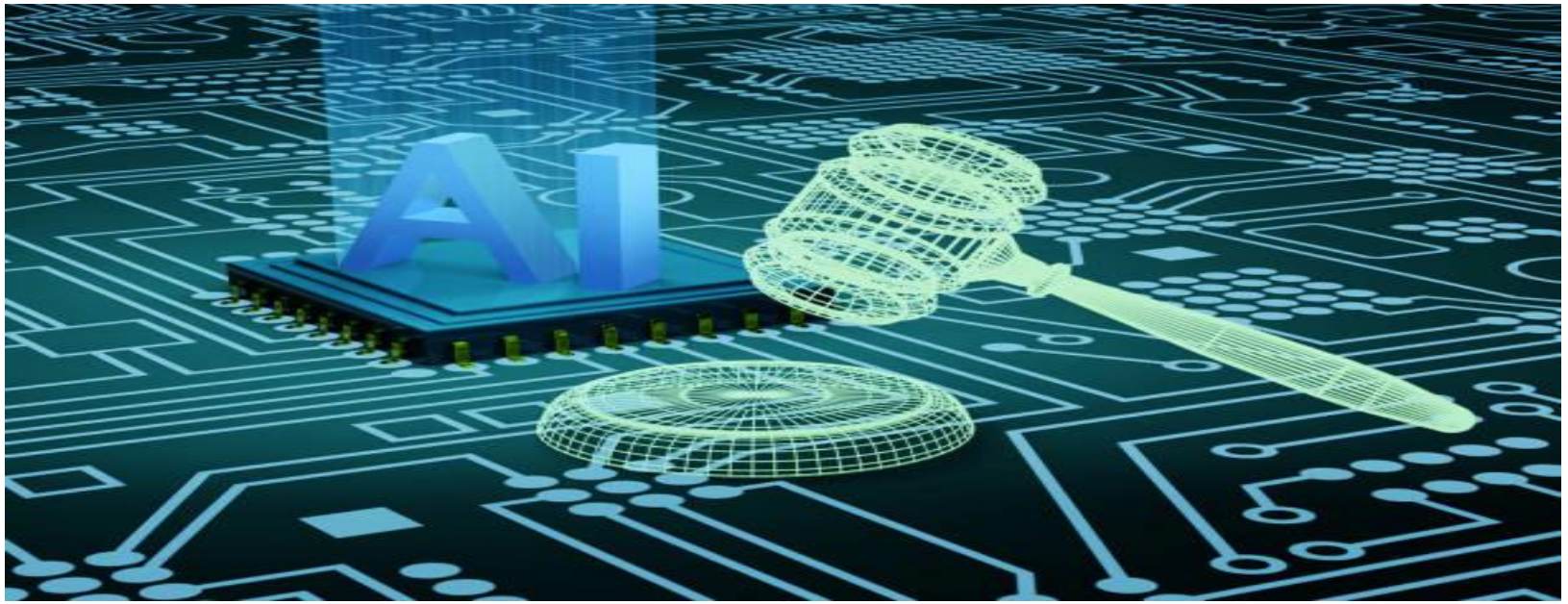
How AI Helps in Risk Assessment

- AI analyzes large datasets to detect hidden risks
- Improves decision-making with real-time insights
- Identifies threats before they occur
- Enhances efficiency and predictive accuracy

Risk Assessment and Management

Policies and Regulations

- Global overview of AI Act EU, NATO, USA, UK, China, Russia
- Need for harmonized standards
- Impact on critical infrastructures



AI Could Have Mitigate the Losses for Heathrow Closure March 21, 2025?

Conclusions and Recommendations

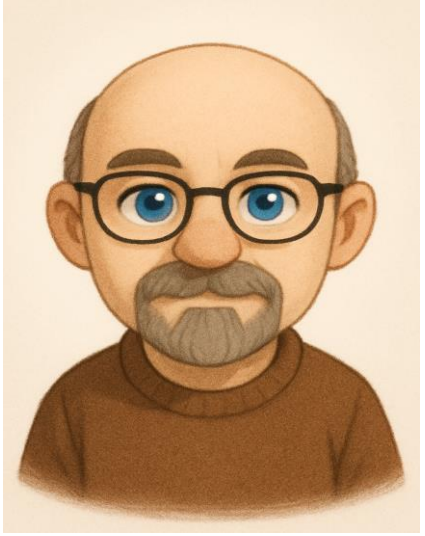
By leveraging AI critical infrastructure operators can enhance resilience, minimize downtime, and mitigate the impact of disruptions, ultimately ensuring the reliable operation of essential services

A combination of AI-based security tools, human expertise and robust security practices should be used to effectively protect critical infrastructure systems against threats of various kinds

Need for governance and regulation


Thank You!

Alberto Stefanini



AIIC – ISACA Rome Chapter

6 Maggio 2025

Applicazioni della Intelligenza Artificiale tra passato, presente e futuro: cosa ci aspettiamo?

Alberto Stefanini

AIIC Associazione Italiana esperti in Infrastrutture Critiche, Rome, Italy

alberto.stefanini@gmail.com

Introduzione

- Intelligenza Artificiale: **paradigma trasformativo** infrastrutture critiche
- **Evoluzione storica:** dai sistemi esperti ai moderni sistemi di Intelligenza Artificiale Generativa (IAG)
- **Obiettivo:** esaminare applicazioni attuali, sfide emergenti e prospettive future

Cos'è l'Intelligenza Artificiale

- **intelligenza artificiale** -> *capacità o tentativo di simulare una generica forma di intelligenza.* Quattro incarnazioni storiche:
  - Albori **informatica** (anni 30-50)
  - Sistemi **rule-based**, sistemi **esperti** (anni 60-90)
  - **Machine learning** (a partire dai 90)
  - Intelligenze artificiali **generative** (oggi)
- **IA in contesti critici** -> *capacità di rispondere a domande complesse e immaginare futuri alternativi*

Prime Applicazioni IA a Infrastrutture Critiche

- **Sistemi esperti**:
  - gestione impianti termici (PROP)
  - Algoritmi di ottimizzazione per la gestione delle reti elettriche
- **Machine learning**:
  - previsione del traffico/gestione delle emergenze

Prime Applicazioni IA: PROP

- Sistema **ibrido** per la gestione di un impianto termico
  - **governo**: rete di Petri per la gestione degli stati principali del sistema (avvio, gestione ordinaria, emergenza, ripristino ecc.)
  - gestione **singoli stati** -> sistemi esperti = set di regole di gestione

Cfr: Gallanti, M., Guida, G., Spampinato, L., and A. Stefanini (1985). **Representing Procedural Knowledge in Expert Systems: an Application to Process Control**. *9th Int. Joint Conference on Artificial Intelligence*, Los Angeles, Aug. 1985.

Cybersecurity Infrastrutture Critiche

- Predizione guasti/manutenzione predittiva: analisi vibrazioni e temperatura treni alta velocità ERTMS – European Rail Traffic Management System
- AI per l'ottimizzazione del traffico urbano e ferroviario AI4Cities
- Smart grid: bilanciamento dinamico domanda-offerta Horizon 2020 FlexGrid
- IA per rilevazione precoce incendi boschivi/terremoti FuegoRed
- Riferirsi al documento del GdL per avere una **rassegna ampia e dettagliata**

Cybersecurity Infrastrutture Critiche (ct.)

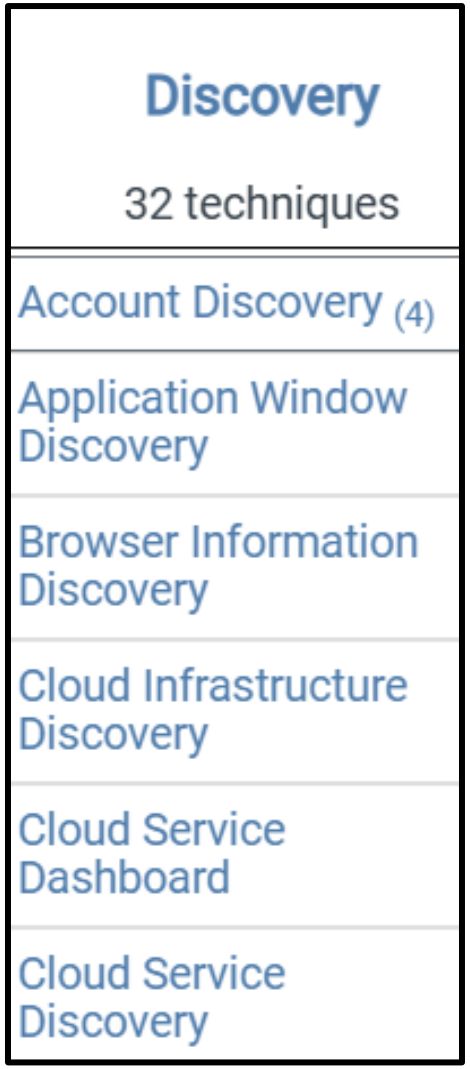
- Sistemi di rilevamento delle intrusioni basati su AI: Darktrace AI
- Automazione nella risposta agli incidenti e mitigazione delle minacce. Deliverables | AI4CYBER
- Integrazione con droni per il monitoraggio post-disastro DJI Disaster Response Program
- Cybersecurity sistemi SCADA controllo reti energetiche NIST Cybersecurity Framework
- Analisi delle vulnerabilità con modelli predittivi: MITRE ATT&CK

NIST - National Institute of Standards and Technology

- Agenzia federale statunitense sviluppo standard/linee guida innovazione e sicurezza
- **NIST Cybersecurity Framework:** strumento volontario gestione rischio cyber
- **Funzioni chiave**:
  - **Identify** – contesto/gestione asset
  - **Protect -** salvaguardie limitazione impatti
  - **Detect** – Rilevare eventi anomali/incidenti
  - **Respond** – Risposta incidenti
  - **Recover** – Ripristino capacità/servizi

MITRE ATT&CK

- A metà fra un glossario ontologico ragionato e un *repository* di tecniche
- Tematiche di riferimento:
  - **Reconnaissance, Resource Development, Initial Access, Execution, Persistence, Privilege Escalation, Defense Evasion, Credential Access, Discovery**
- Ogni *entry* ha almeno una decina di riferimenti afferenti
- Ad es. la lista: **Discovery** riporta **32 tecniche diverse**



ENISA: foresight cybersicurezza

- *Foresight:* obiettivo strategico principale ENISA: **riflettere** su scenari futuri e **prepararsi** strategicamente
- metodologia di foresight ENISA:
  - Identificare minacce emergenti alla cybersicurezza
  - Mappare sfide future
  - Dare priorità a sforzi su tecnologie emergenti

Normativa UE armonizzazione cybersicurezza IA

- **Regolamento europeo Intelligenza Artificiale**
  - AI Act – Reg. UE 2024/1689 in vigore 1° agosto 2024
  - Primo quadro normativo completo IA
  - Approccio basato su rischio/sicurezza prodotto
- **Armonizzazione standard** cybersicurezza:
  - Mappatura terminologica tra ISO 27001, IEC 62443, NERC CIP
  - Analisi semantica automatica definizioni/prescrizioni
  - Supporto PMI e auditor
  - Coerente prospettive **ENISA** IA/governance rischio

Sfide Attuali

- **Affidabilità** algoritmi
  - E.g., riduzione falsi positivi in analisi sicurezza
  - In genere la c.d. hallucination
- **Scalabilità** IA a infrastrutture complesse
- **Interoperabilità** standard di cybersecurity (ISO 27001, IEC 62443, NIST CSF)
  - ENISA AI Cybersecurity

Opportunità Future e Nuove Frontiere

- **Gemelli digitali** per la manutenzione predittiva avanzata Siemens Digital Twin
- **IA e reti 6G** per l'analisi predittiva su larga scala Hexa-X
- **IA per la resilienza urbana**: monitoraggio e risposta adattiva a eventi critici RESILOC
- **GALICIA**: Generative AI with Cybersecurity for Internet Applications development, https://www.galicia-project.eu/project/

Siemens Digital Twin - Replica virtuale asset fisici/ processi

- **Obiettivi e Vantaggi**
  - Migliorare efficienza e sostenibilità
  - Ridurre costi e tempi di sviluppo
  - Mitigare rischi e prevedere criticità
  - Supportare la trasformazione digitale delle imprese
- **Componenti Offerti da Siemens**
  - Software simulazione e progettazione
  - Soluzioni automazione industriale
  - Servizi integrazione/supporto

Siemens Digital Twin - Replica virtuale asset fisici/ processi

Settori Applicativi Principali

- **Automotive:** progettazione veicoli, test di sicurezza, guida autonoma
- **Manifatturiero:** ottimizzazione processo produzione
- **Energia:** simulazione reti intelligenti/impianti
- **Sanità:** progettazione dispositivi medici/digitalizzazione ospedaliera
- **Infrastrutture:** edifici intelligenti, città digitali

Vedi: Siemens Search

GALICIA: metodo e piattaforma

- GALICIA (https://www.galicia-project.eu/project/):
  - metodi formali verifica correttezza/sicurezza codice generato da IA
  - conformità requisiti utenti e standard settore
- **Piattaforma GALICIA:**
  - interagire con **diverse sorgenti** di intelligenza artificiale generativa
  - **produrre** codice sorgente e **verificarlo** tramite diversi modelli formali
  - Utilizzo risultati: convalidare **corrispondenza** a intento originale e sicurezza operativa
  - **Fuochi d'interesse** correnti: *cybersecurity, stress testing, reverse engineering, user trust, code quality*

Conclusioni e Prospettive di Ricerca

- IA continuerà a trasformare la gestione delle infrastrutture critiche
- Necessità di coniugare innovazione tecnologica con regolamentazione e *governance*
- Opportunità per nuove collaborazioni e progetti di ricerca nel monitoraggio reti e cybersecurity *AI-driven*

Applicazioni della Intelligenza Artificiale tra passato, presente e futuro: cosa ci aspettiamo?

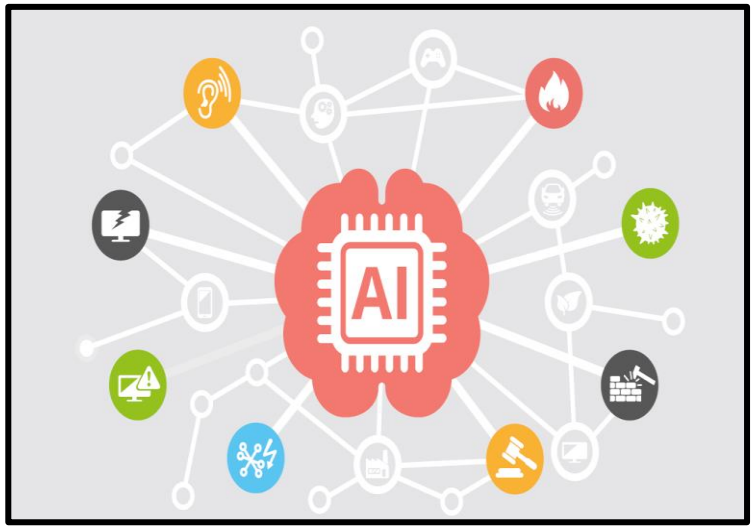
alberto.stefanini@gmail.com

IA e impatto ambientale

- L'Intelligenza Artificiale Generative oggetto di critiche per il suo **impatto ambientale**
- il **consumo energetico** per addestrare modelli LLM comparabile a quello di una piccola città:
  - i processi di training avvengono una tantum
  - l'utilizzo successivo dei modelli ottimizzati è **molto più efficiente** dal punto di vista energetico
- i benefici offerti dai LLM sono notevoli:
  - accelerazione della **ricerca scientifica**
  - drastica riduzione dei tempi di redazione di **testi specialistici**
  - ottimizzazione di processi **produttivi e decisionali**
  - sostituzione di parte **comunicazione sincrona** (riunioni, viaggi di lavoro) con interazioni **asincrone** ed efficienti via AI

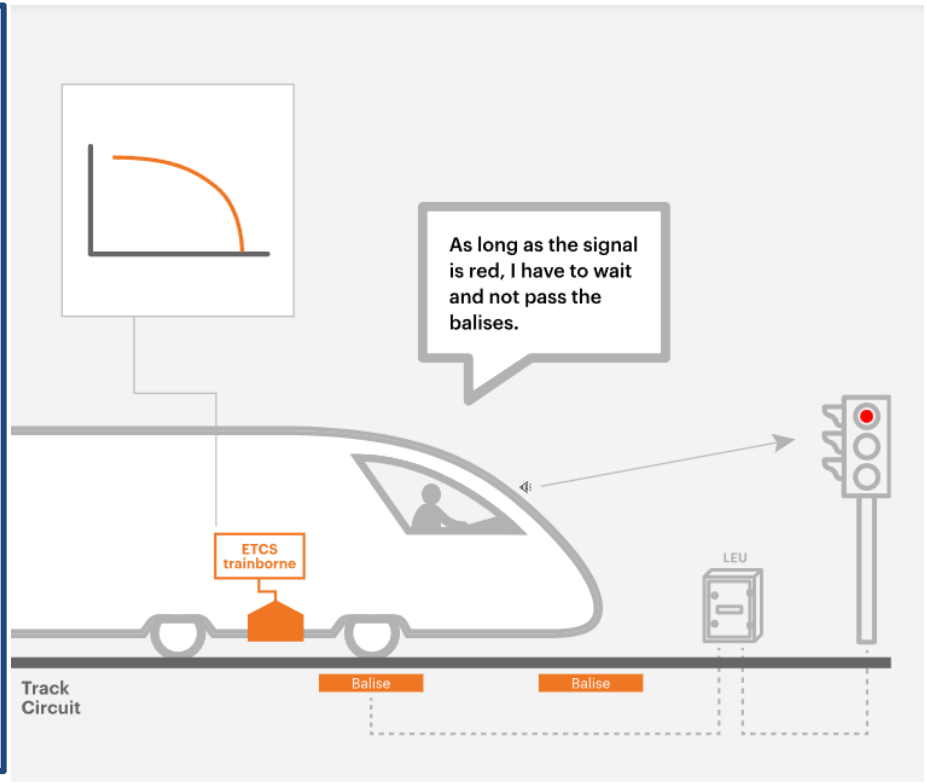
ENISA e foresight cybersicurezza (ct.)

- studi e rapporti su tecnologie critiche:
  - AI e cybersicurezza (2023–2024)
  - Cyber insurance e AI
- Affidabilità, standardizzazione e privacy - IA sfida e risorsa
  - attaccare o manipolare
  - rafforzare le difese e le capacità di risposta



European Rail Traffic Management System - ERTMS

- **ERTMS/ETCS Level 1**
  - Eurobalise without infill
  - Overlay to Existing Signaling System
  - Movement Authorities through Eurobalise
  - Train Integrity & Position by Track Circuit



A balise is an electronic beacon or transponder placed between the rails of a railway as part of an automatic train protection (ATP) system.
The French word balise is used to distinguish these beacons from other kinds of beacons

Luigi Carrozzi

Critical Infrastructure Resilience and Artificial Intelligence

Webinar a cura AIIC & ISACA Rome Chapter

06 Maggio 2025, ore 15:00

Vincoli Etici e Sociali nell'utilizzo della Intelligenza Artificiale nelle Infrastrutture Critiche: sono soltanto impedimenti?

Impatti organizzativi nell'adozione dei sistemi di Intelligenza Artificiale

Luigi Carrozzi

(CGEIT, CRISC, ISMS Auditor L.A., ISMS Senior Manager, Privacy Officer)

EMERGING TECHNOLOGY RISK ASSESSMENT: *THE POLICY, LEGAL, ETHICAL, AND REGULATORY* FACTOR FOR TECH. AVAILABILTY

ARTIFICIAL INTELLIGENCE TECHNOLOGY AVAILABILITY FOR CRITICAL INFRASTRUCTURE APPLICATIONS

Scenario	Science and Technology Maturity	Use Case, Demand, and Market Forces	Resources Committed	Policy, Legal, Ethical, and Regulatory Impediments	Technology Accessibility	T_{AV} Average
Short term (up to 3 years)	2	2.5	3	5	2	2.9
Medium term (3–5 years)	2.5	3	4	5	2.5	3.4
Long term (5–10 years)	3	3.5	5	4	3	3.7

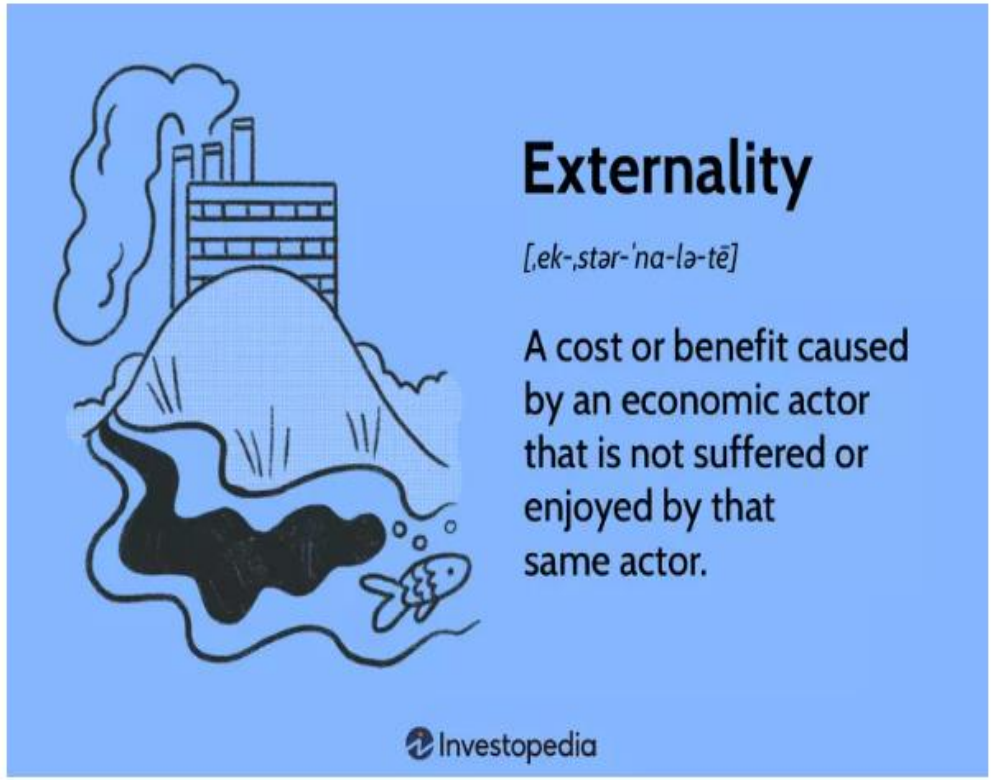
NOTE: The emerging technology risk assessment scale is 0 to < 2 = low impact or not likely feasible, 2 to < 4 = moderate impact or possible, and 4 to 5 = high impact or likely feasible.

Fonte: US HOMELAND SECURITY OPERATIONAL ANALYSIS CENTER – RESEARCH REPORT «EMERGING TECHNOLOGY AND RISK ANALYSIS: ARTIFICIAL INTELLIGENCE AND CRITICAL INFRASTRUCTURE»

POLICY, LEGAL, ETHICAL AND REGULATORY CONSTRAINS

ARE ONLY IMPEDIMENTS?

«EXTERNALITY»



Investopedia / Madelyn Goodnight

Source: https://www.investopedia.com/

EXAMPLE OF NEGATIVE EXTERNALITY

A company pollutes the environment during the production process

The cost of pollution is not borne by the factory, but is imposed on the entire society

BRUCE SCHEINER AND THE "NEGATIVE EXTERANLITIES" OF INSECURE SOFTWARE - OCTOBER 2004 -

Schneier on Security

Blog | Newsletter | Books | **Essays** | News | Talks | Academic | About Me

Home > Essays

Information Security and Externalities

Bruce Schneier
ENISA (European Network and Information Security Agency) Quarterly
January 2007

This essay is an update of Information security: How liable should vendors be?, Computerworld, October 28, 2004.

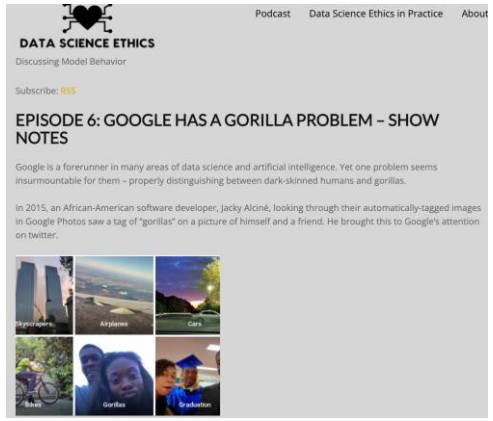
Bruce Sheneir
Public Interst Technologist

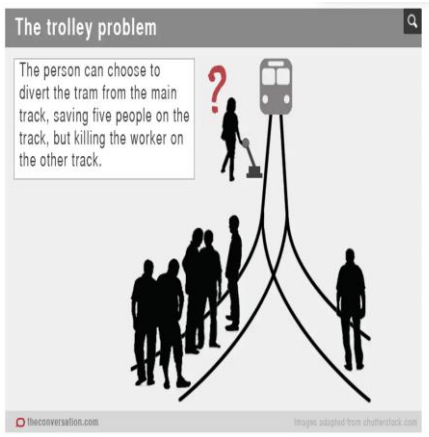
…insecure software It is a result of bad design, poorly implemented features, inadequate testing and security vulnerabilities from software bugs.

…The money we spend on security is to deal with the myriad effects of insecure software….

Source: Information Security and Externalities - Schneier on Security

AI ETHICAL CONCERNS





NEURALINK

FONTE: https://neuralink.com

A multi-dimensional approach to disinformation

Report of the independent High level Group on fake news and online disinformation

ARTIFICIAL INTELLIGENCE HARMS

INDIVIDUAL HARMS

- Physical or mental harm
- Loss or restrictions of Rights/Freedoms
- Restriction to the access of goods and services
- Discrimination
- Unfairness
- Identify theft/loss of identity
- Disclosure of personal information
- Damage to reputation
- Monetary loss
- ……

COLLECTIVE /SOCIETAL HARMS

- Democracy threats
- Fake news/Disinformation
- Group privacy violations
- Collapse of essential services
- Freedom of opinion limitations
- Discrimination
- ……

The need of a AI governance
that balance
business objectives with
Individual and public benefit objectives

REGULATION
ON THE PROTECTION OF NATURAL PERSONS
WITH REGARD TO THE PROCESSING OF PERSONAL DATA
AND ON THE FREE MOVEMENT OF SUCH DATA

Recital 7) "***Those [technology] developments require a strong and more coherent data protection framework*** *in the Union, backed by strong enforcement,* ***given the importance of creating the trust that will allow the digital economy to develop across the internal market"***

REGULATION
LAYING DOWN HARMONISED RULES
ON ARTIFICIAL INTELLIGENCE..

Art.1.1) *"The purpose of this Regulation is* ***to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence*** *(AI), while* ***ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation"***

AI HLEG - Trustworthy Artificial Intelligence

Principles

- **RESPECT FOR HUMAN AUTONOMY**
- **PREVENTION OF HARM**
- **FAIRNESS**
- **EXPLICABILITY**

Requirements

1 **Human agency and oversight**

Including fundamental rights, human agency and human oversight

2 **Technical robustness and safety**

Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility

3 **Privacy and data governance**

Including respect for privacy, quality and integrity of data, and access to data

4 **Transparency**

Including traceability, explainability and communication

5 **Diversity, non-discrimination and fairness**

Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation

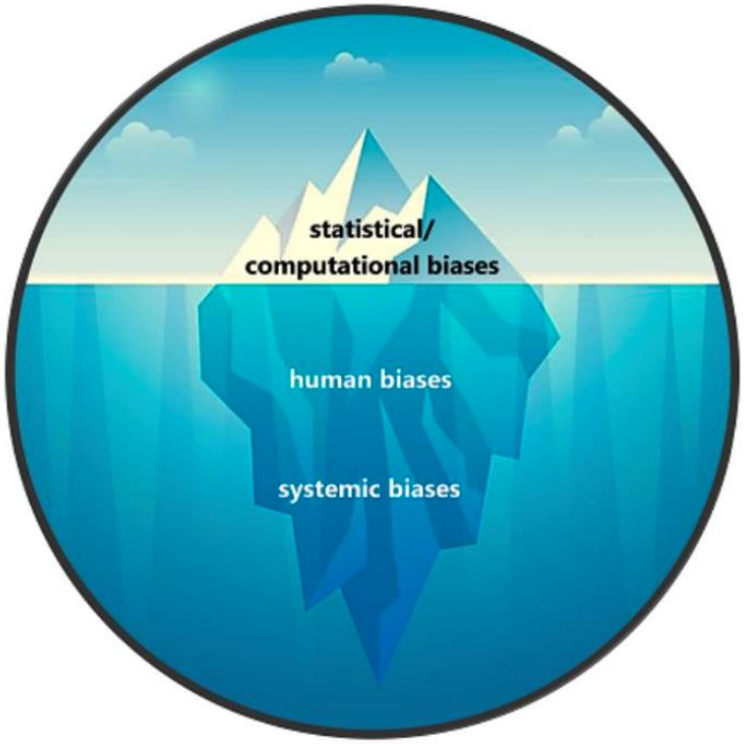
6 **Societal and environmental wellbeing**

Including sustainability and environmental friendliness, social impact, society and democracy

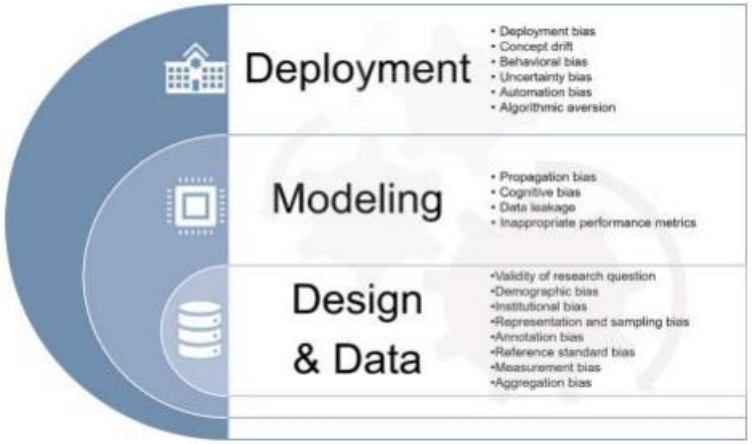
7 **Accountability**

Including auditability, minimisation and reporting of negative impact, trade-offs and redress.

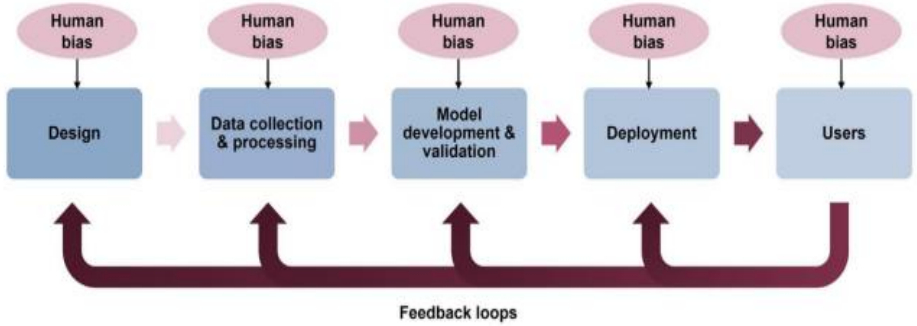
Organizational impacts of AI system adoption
One for all: the necessity to manage AI BIAS



The challenge of managing AI bias
*(Source: National Institute of Standards and Technology. (2022). **NIST Special Publication 1270** - Towards a Standard for Identifying and Managing Bias in Artificial Intelligence)*



Main types and sources of bias



Human bias in the artificial intelligence life cycle

(Source: *Koçak B, Ponsiglione A, Stanzione A, et al. (2024). "Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects".* Diagnostic and Interventional Radiology. DOI: 10.4274/dir.2024.242854)

OUTCOME Vs. *PROCESS* THINKING

«Outcome» Vs. **«Process»**

- Focused on the **outcome**
- **The way in which the result is achieved,** i.e. the **functioning of the technology**, **is secondary**

Focuses the **structure** (components) of the AI system **and its functions**

Particular **attention is given to process:** criteria for **input data** and **algorithms.**

It enables the **identification** of the AI systems that are based on **"incorrect" processes**

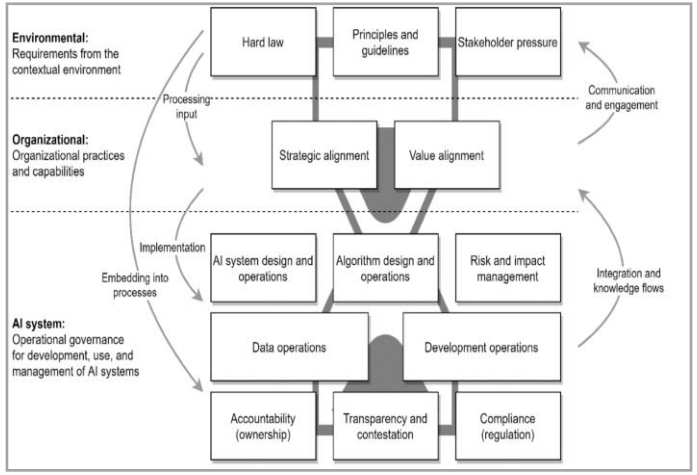
Cfr.: Giulia Gentile, Pensare all'intelligenza artificiale: "outcome thinking" vs "process thinking", due visioni a confrontoPollicino & Partners Advisory, Università dell'Essex. - Il sole 24 ore NT+diritto. L'autrice nell'articolo cita il volume di Richard Susskind, "How to Think About AI: A Guide for the Perplexed"

ORGANIZATIONS AND THE CHALLANGE OF AI TRUSTWORTHINESS INTO PRACTICE

INTERNATIONAL STANDARD **ISO/IEC 38507:2022**

Edition 1
2022-04

Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations



The hourglass model of organizational AI governance
(Source: Matti Mäntymäki, Matti Minkkinen, Teemu Birkstedt, Mika Viljanen. (2023). Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance - arXiv:2206.00335 [cs.AI])

INTERNATIONAL STANDARD **ISO/IEC 42001:2023**

Edition 1
2023-12

Information technology — Artificial intelligence — Management system

NIST Artificial Intelligence Risk Management Framework

INTERNATIONAL STANDARD **ISO/IEC 23894:2023**

Edition 1
2023-02

Information technology — Artificial intelligence — Guidance on risk management

THE «REDUCED RISK ON INVESTMENT» OF AI TRUSTWORTHINESS

$$\frac{V_T}{C_T} \gg 1$$

…the value generated by investing in AI Trustworthiness may result much greater than the cost required to achieve it

TRUSTWORTHINESS PAYS!

Francesca Della Mea

Come affrontare il tema della
valutazione del rischio delle infrastrutture critiche
che fanno uso dell'Intelligenza Artificiale:
possiamo farne a meno?

Francesca Della Mea

6 Maggio 2025

Introduzione

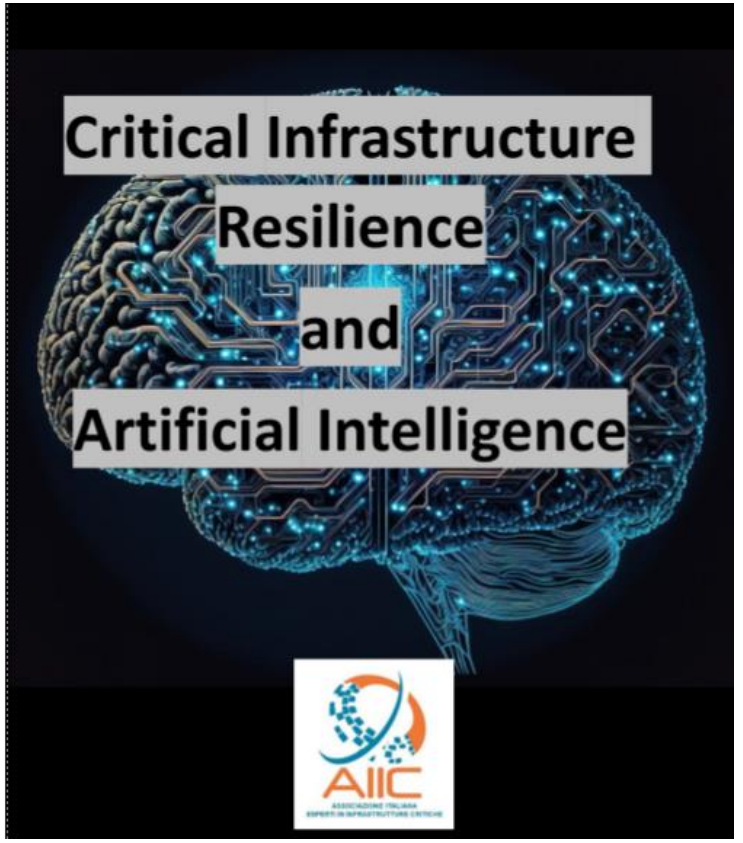
Questo approfondimento:

Come affrontare il tema della valutazione del rischio delle infrastrutture critiche che fanno uso dell'Intelligenza Artificiale: possiamo farne a meno?

Discute alcuni dei temi presentati in:

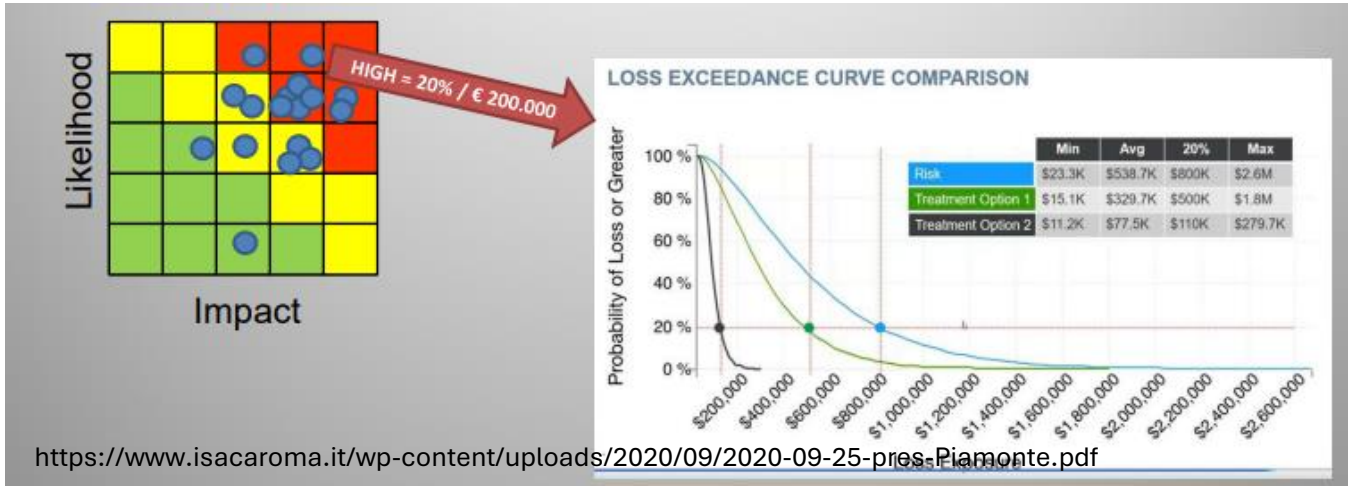
«RISK ASSESSMENT AND MANAGEMENT» di Glauco Bertocchi, Francesca Della Mea, Giorgio Pizzi

in «Critical Infrastructure Resilience and Artificial Intelligence», Chap.5



Il Risk assessment è un’attività complessa

- Complessità degli elementi da valutare (tecnici, legali, organizzativi. umani, …)
- Nel caso delle Infrastrutture Critiche gli impatti possono estendersi alla società e alla sicurezza delle persone



https://www.isacaroma.it/wp-content/uploads/2020/09/2020-09-25-pres-Piamonte.pdf

Ruolo e autonomia dell’analista del rischio AI:

- *Raccolta dei dati: supportata da strumenti (AI-based) o no?*
- *Su quali dati (storico-esperienziali) basarsi)?*

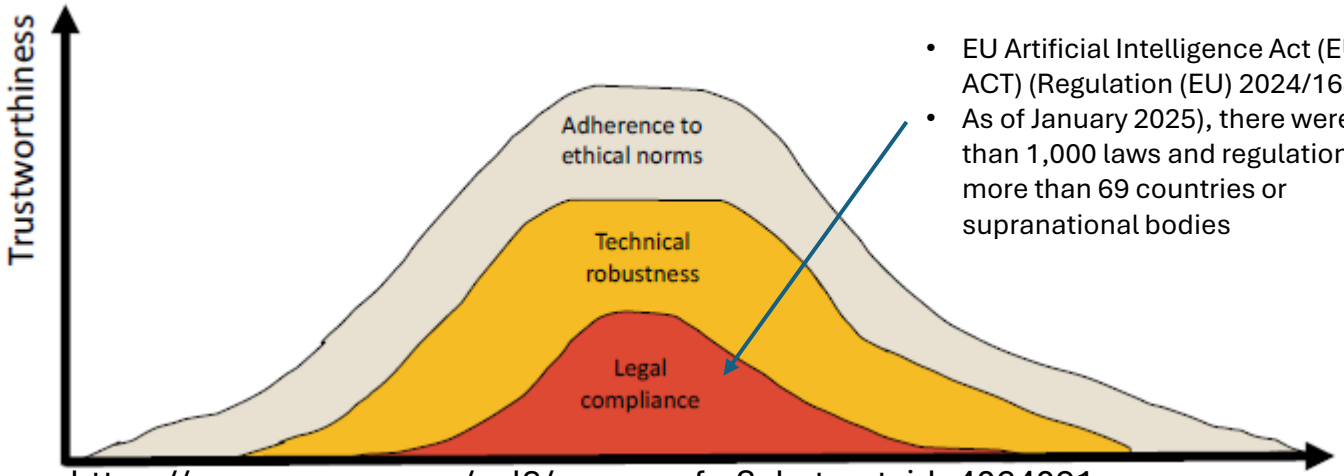
Quali rischi?

- Non solo cybersecurity (CIA / CIAA), ma trustworthy AI / ethics

Trusthworthy AI:

- valid and **reliable, safe, secure and resilient**, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful biases managed

- Resistenza agli attacchi e robustezza delle infrastrutture
- Valutazioni di business e governance (etica)
- Interazioni con la componente umana (safe, explainable, privacy-enhanced, …)



https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091

- EU Artificial Intelligence Act (EU AI ACT) (Regulation (EU) 2024/1689)98
- As of January 2025), there were more than 1,000 laws and regulations in more than 69 countries or supranational bodies

- *Team di valutazione del rischio strutturati e multidisciplinari*
- *Ruoli?*

Quali scenari di utilizzo nelle CI?

- I rischi devono essere valutati **nel contesto**, facendo riferimento allo **scenario di utilizzo**

  - Esempi di utilizzo dell'AI nelle CI*
    - Operational awareness – ad es. monitoraggio del traffico di rete, identificazione attività non usuali
    - Ottimizzazione delle Performance – dei sistemi IT, ma anche della supply chain
    - Automazione delle operazioni - esecuzioni di compiti di routine, analisi di grandi volumi di dati
    - Event Detection:
    - Forecasting
    - System planning
    - Customer service automation
    - Modeling & Simulation
    - Physical security (survelliance data analysis)
    - …

- *Rilevanza di contesto, scenario*

*DHS - MITIGATING ARTIFICIAL INTELLIGENCE (AI) RISK - Safety and Security Guidelines for Critical Infrastructure Owners and Operators

Quali strumenti possiamo adottare?

NIST AI Risk Management Framework

- La proprietà del Sistema AI (*valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed*) sono valutate in relazione con il contesto in cui il Sistema sarà adottato
- Articolato attorno a funzioni CORE (Govern, Map, Measure, Manage)
- Con documenti a supporto (NISTIR 831283 ¸ NISTIR 836784, NIST SP 1270 85)

ISO/IEC 42001:2023 Artificial intelligence —Management system)

- Anche: ISO/IEC 23894:2023 (Information technology — Artificial intelligence — Guidance on risk management)
- In sviluppo ISO/IEC CD 27090 (Cybersecurity — Artificial Intelligence — Guidance for addressing security threats and failures in artificial intelligence systems)

Strumenti: framework, standard, best practice

Strumenti: ENISA e CISA

ENISA - Multilayer Framework for Good Cybersecurity Practices for AI

https://www.enisa.europa.eu/publications/multilayer-framework-for-good-cybersecurity-practices-for-ai

Il report ENISA propone un approccio a layer per migliorare la cybersecurity delle infrastrutture ICT con componenti AI. L'approccio è basato su risk management.

NSA - Joint Guidance on Deploying AI Systems Securely

- Pubblicato da NSA congiuntamente con agenzie di Cybersecurity di USA, Canada, Australia, New Zealand, e UK (incluse CISA, FBI, DHS)
- Come indicato dal sottotitolo "Best Practices for Deploying Secure and Resilient AI Systems", il documento include referenze ed è una guida pratica ed autorevole.

Strumenti specifici su Infrastrutture Critiche

MITRE - Principles for Reducing AI Cyber Risk in Critical Infrastructure: A Prioritization Approach

Basato sul livello di rischio delle funzioni di infrastrutture critiche basate su AI.

Prima di decidere se

- Applicare l'AI a funzioni infrastrutturali critiche
- Dove applicarla

È necessario definire qual è il livello di possibili conseguenze non accettabili.

US Homeland Security - MITIGATING ARTIFICIAL INTELLIGENCE (AI) RISK: Safety and Security Guidelines for Critical Infrastructure Owners and Operators

Le linee guida sono mappate su NIST AI RMF.

Un documento sintetico che "specifically addresses risks to safety and security which are uniquely consequential to critical infrastructure."

Tassonomie di rischi AI

Per la valutazione dei rischi, visto lo storico limitato, può essere utilizzato un approccio analitico, basato su tassonomie dettagliate dei rischi AI

ENISA - Cybersecurity Challenges. Threat Landscape for Artificial Intelligence.

Contiene due ampie tassonomie, collegate rispettivamente agli asset e alle minacce.
Le minacce sono raggruppate in:

- Nefarious Activity/Abuse, Unintentional Damage, Legal, Failures or malfunctions, Eavesdropping Interception Hijacking, Physical attacks, Outages, Disasters

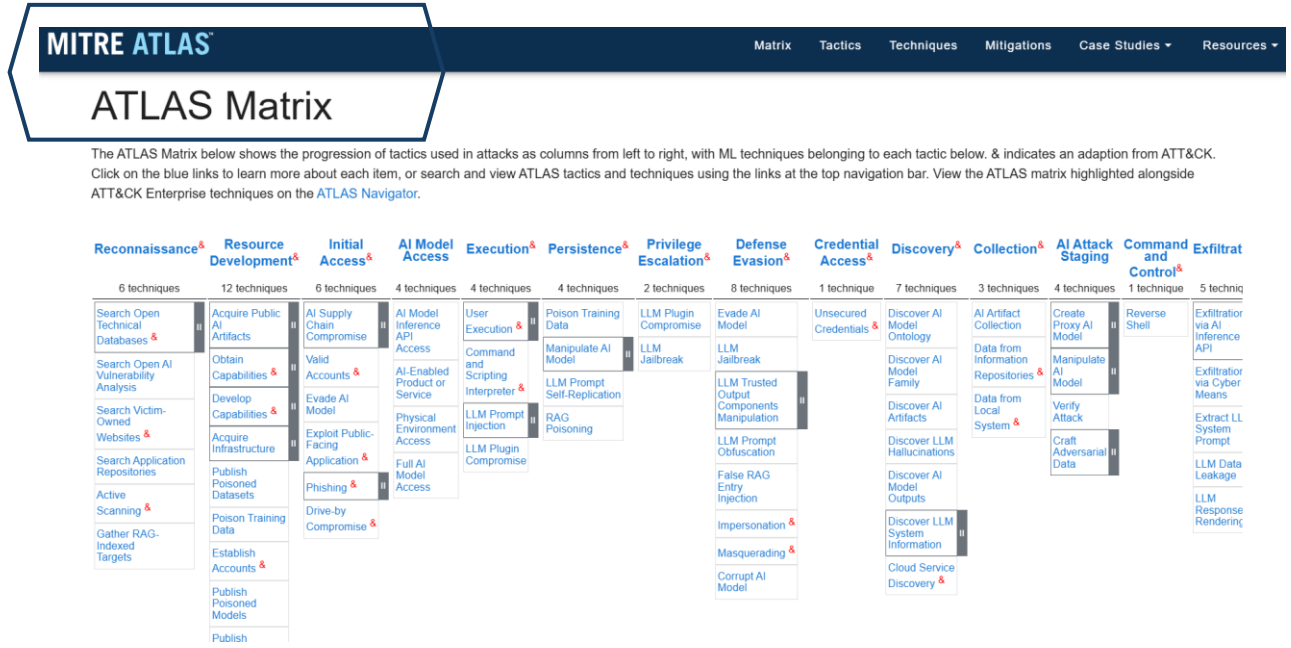
ENISA - "Securing Machine Learning Algorithms

Minacce e vulnerabilità che si applicano specificamente agli algoritmi di ML, mappate poi rispetto ai controlli dei principali framework disponibili (2021)

Tassonomie di rischi AI

MITRE ATLAS
(Adversarial Threat Landscape for AI Systems)

- Knowledge base of adversary tactics and techniques against AI-enabled systems
- Alcune delle tecniche sono specficihe di sistemi AI enabled (LLM Prompt Injection, Evade ML model, Denial of ML Service, Erode ML model or dataset integrity)



OWASP - Top 10 for Large Language Model Applications

https://owasp.org/www-project-top-10-for-large-language-model-applications/

Le 10 minuacce più rilevanti per i LLM, con dettagliate vulnerabilità e strategie di mitigazione e mitigazione, esempi, referenze.

Tassonomie di rischi AI – CI

DHS - MITIGATING ARTIFICIAL INTELLIGENCE (AI) RISK - Safety and Security Guidelines for Critical Infrastructure Owners and Operators

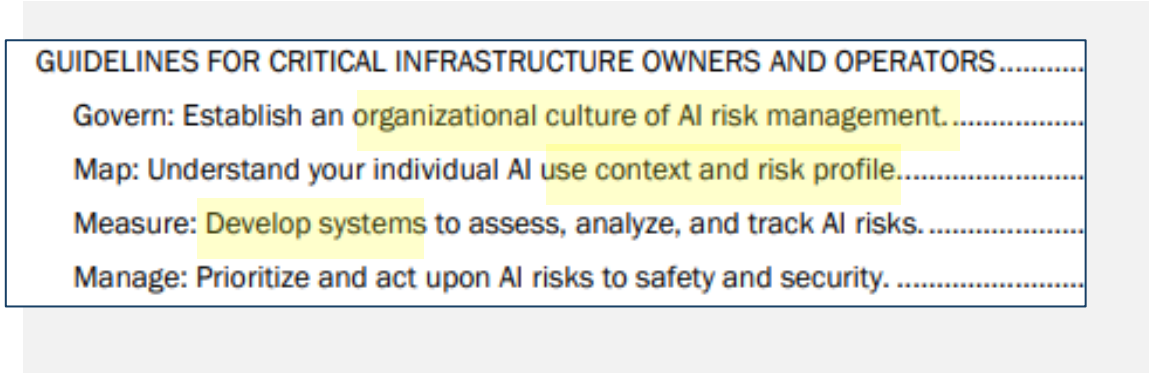
Descrive sia le linee guida per gli operatori delle infrastrutture critiche, ed elenca le principali categorie di rischio, individuando delle sotto-categorie:

Risk Category: Attacks on AI

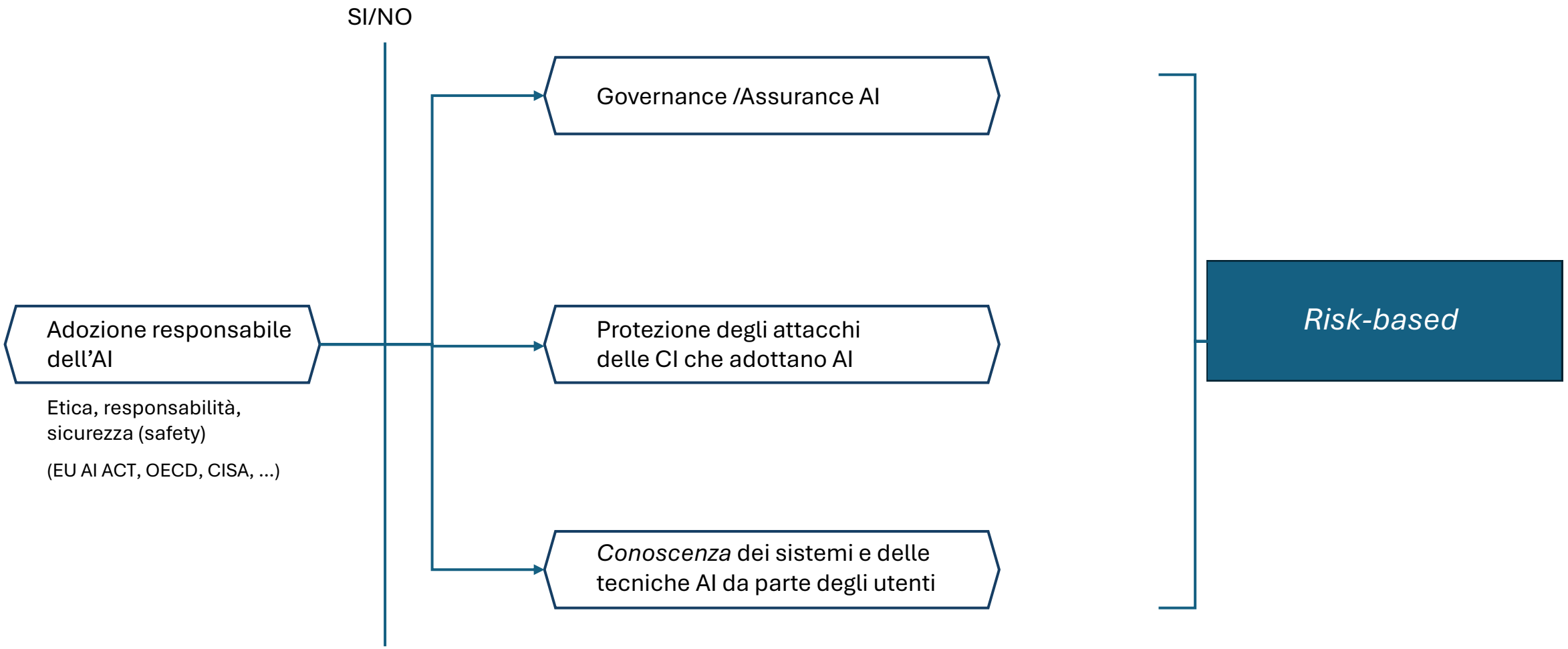
- Adversarial Manipulation of AI Algorithms or Data
- Evasion Attacks (The malicious injection of prompts into an AI system to bypass the model)
- Interruption of Service Attacks
- Loss of Data
- Model Inversion and Extraction

Risk Category: AI Design and Implementation Failures

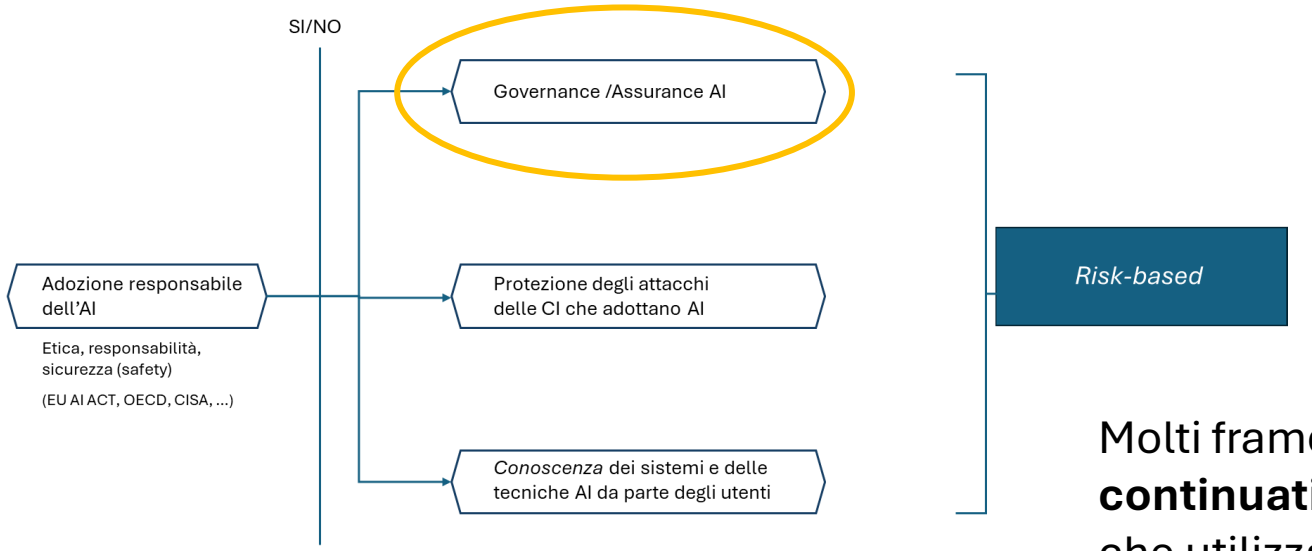
…



In sintesi


SI/NO
Governance /Assurance AI
Adozione responsabile dell'AI
Protezione degli attacchi delle CI che adottano AI
Etica, responsabilità, sicurezza (safety)
(EU AI ACT, OECD, CISA, …)
Conoscenza dei sistemi e delle tecniche AI da parte degli utenti
Risk-based

In sintesi – Governance / Assurance

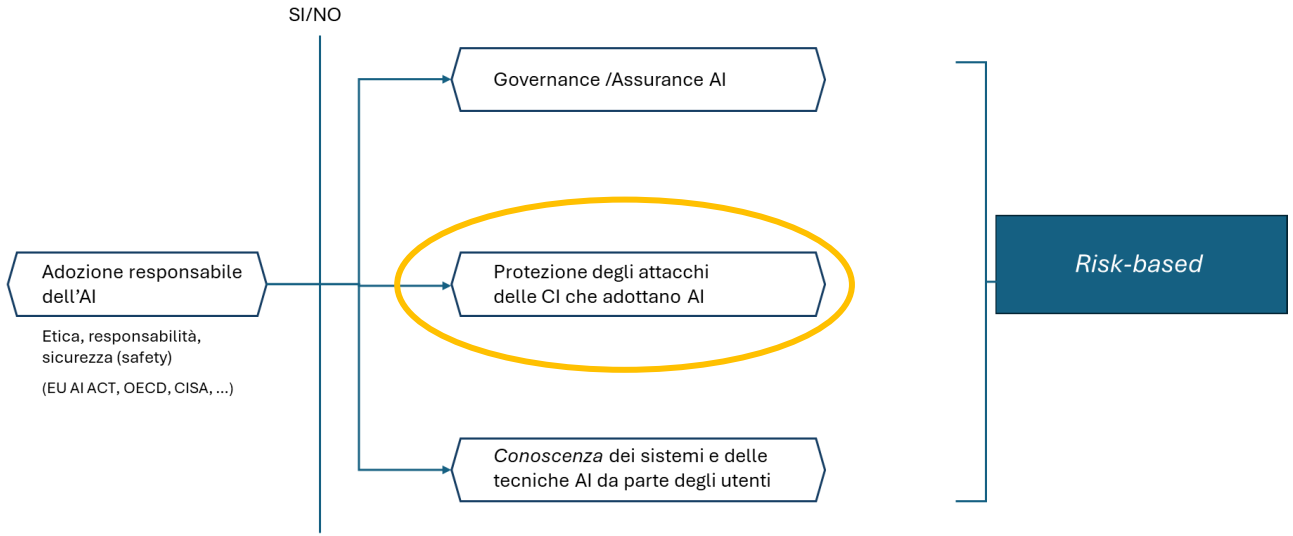


Molti framework si concentrano su **processi continuative** durante il ciclo di vita di un sistema che utilizza AI:

- AI System Assurance (MITRE, CISA, capAI, ...)
- Management Systems – Governance (ISO, ...)

Attenzione anche ai processi correlati, quali ad esempio la Data Governance

In sintesi – Protezione dagli attacchi

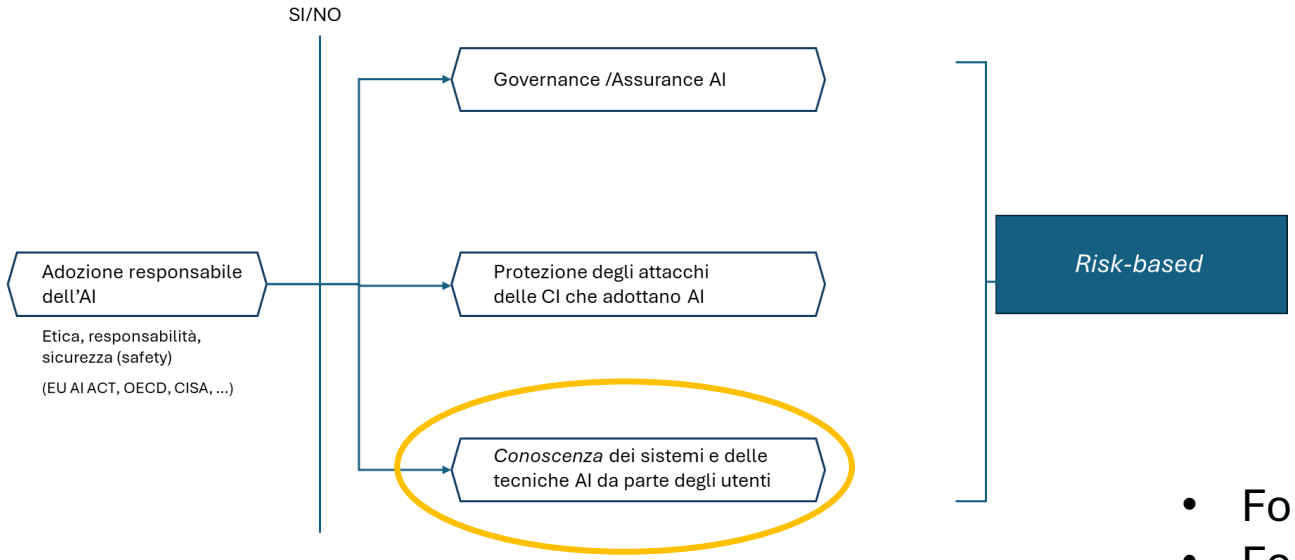


- Alcuni strumenti si focalizzano sull'individuazione e la difesa da attacchi alle applicazione AI based. Elencano le minacce, i possibili impatti, le contromisure

Le possibili contromisure non riguardano solo le tecnologie. CISA e DHS, che si focalizzano su CI, identificano tra gli approcci di mitigazione:

- Ridondanza dei processi, prevedendo l'alternativa tra operatività guidata da AI e "manuale"
- Riduzione del valore dei dati per l'attaccante (tramite Data Masking o ridisegno mirato)
- Validazione dei dataset: impegno addizionale nella protezione dei dataset di training

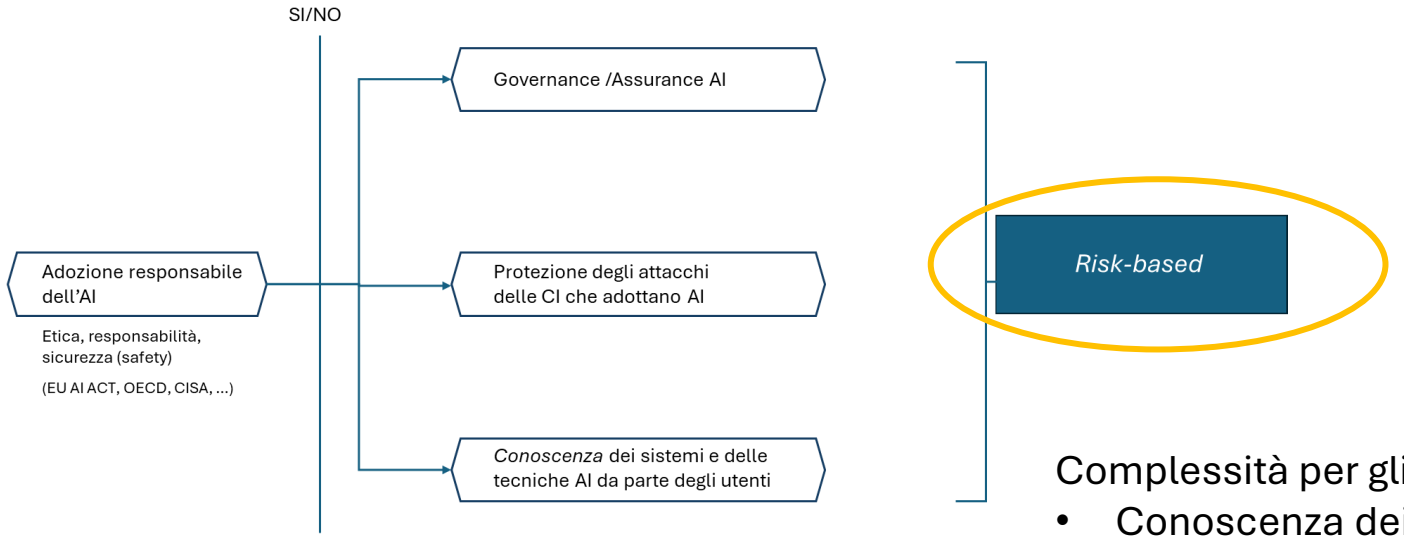
In sintesi – Preparazione degli utenti



- Formazione agli utenti sull'uso dell'AI (generale)
- Formazione agli utenti che utilizzano sistemi AI in processi critici
- Simulazioni di crisi, esercitazioni

Gestire l'introduzione dell'AI come un programma di trasformazione

In sintesi – L'esecuzione dell'analisi



Complessità per gli analisti:

- Conoscenza dei metodi per la valutazione dei rischi vs conoscenze specifiche dei rischi dell'AI
  - Competenze multidisciplinari
- Come superare la mancanza di storico delle valutazioni
- Scarsa visibilità sugli input dell'analisi (controllo)
  - Raccolta dei dati: supportata da strumenti (AI-based) o no? Come verificare i risultati?
  - Molti strumenti (es. explainability) acquisiti da piattaforme di data management o AI
- Documentazione delle assunzioni

In conclusione

- L’introduzione di **sistemi basati su AI,** introduce nuovi rischi, e amplifica rischi esistenti
- E’ fondamentale il governo del ciclo di vita dello strumento sin dalle prime fasi: i rischi potrebbero superare i benefici
- Per la valutazione del rischio delle AI, il contesto è particolarmente rilevante

- L'utilizzo delle tecnologie di **AI nelle CI** introduce opportunità e rischi. Il contesto in cui deve essere effettuata una valutazione del rischio diventa, se possibile, molto più complesso
- Non c’è un unico approccio per valutare i rischi, ci sono tecniche, sempre più complesse, che devono essere applicate alla valutazione. L’analisi deve essere estesa a tutti gli aspetti in perimetro e alle interfacce

- *Paradossalmente*, l’uso delle tecnologie di AI è praticamente indispensabile all’analista del rischio per affrontare queste nuove sfide
- Gli strumenti di per la **valutazione del rischio basati su AI** devono essere governate e difesi adottando gli stessi approcci e framework di tutte gli altri sistemi basati sulla stessa tecnologia

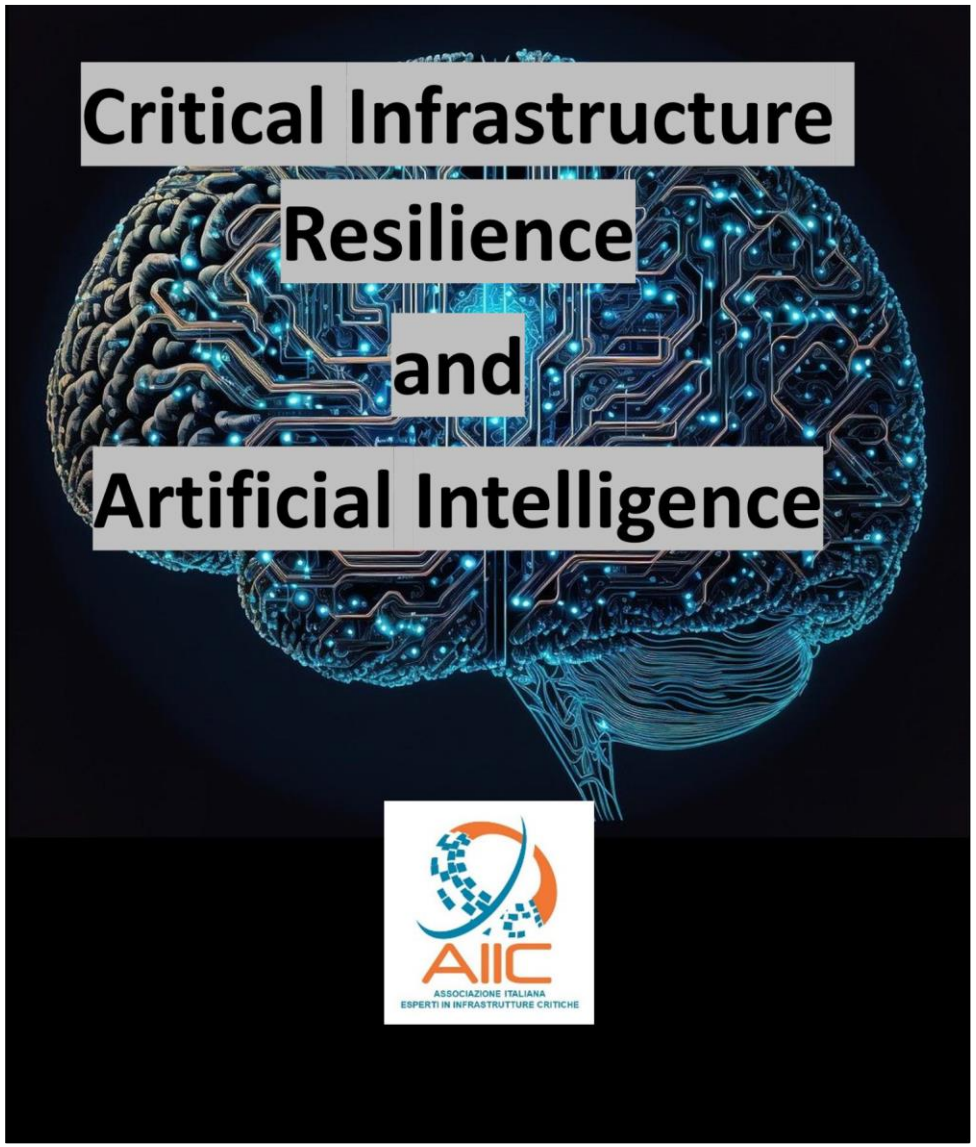
Alberto Caruso de Carolis

AIIC-ISACAROMA 6 maggio 2025 "CRITICAL INFRASTRUCTURE RESILIENCE and ARTIFICIAL INTELLIGENCE"

Vincoli Normativi ed Etici non bastano ad assicurare la Resilienza.

il caso Heathrow e un auspicabile contributo della intelligenza artificiale.

Alberto Caruso de Carolis



2 POLICY AND REGULATIONS: WORLDWIDE OVERVIEW STRATEGIES ON IA & STANDARDS.

(Alberto Caruso de Carolis, Raffaella D'Alessandro, Luisa Franchina, Adriana Peduto, Maria Beatrice Versaci)

2.1 **Overview of current regulations and standards related to the resilience of critical infrastructure and the use of artificial intelligence.**

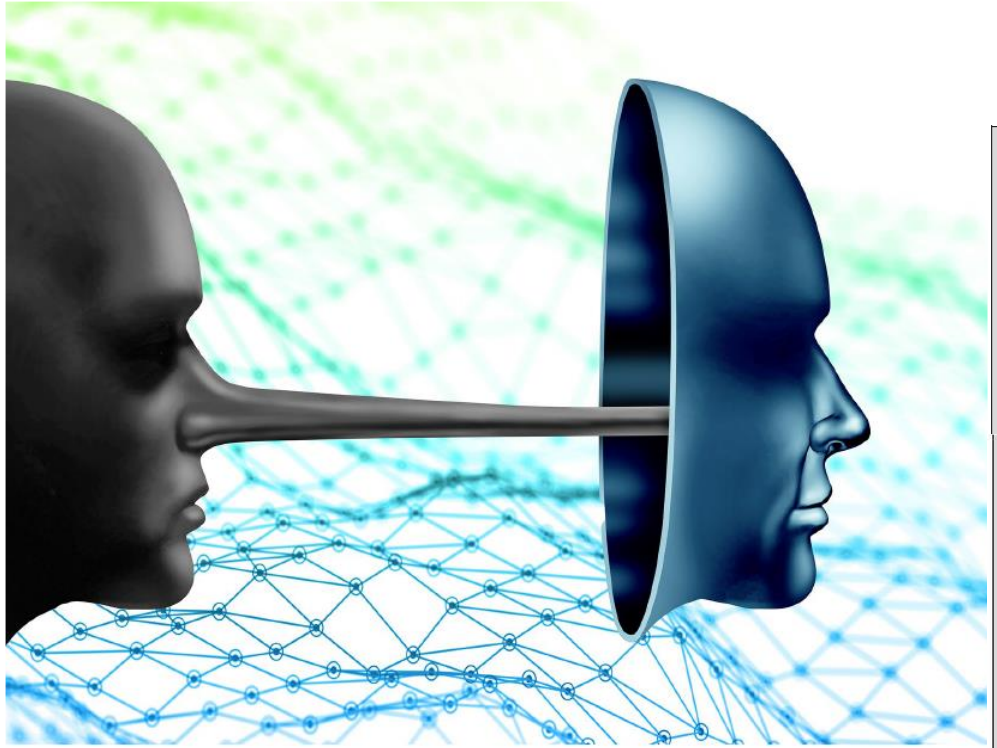
(Luisa Franchina, Maria Beatrice Versaci)

2.2 **Impact of Regulations on AI Adoption**

(Alberto Caruso de Carolis, Adriana Peduto)

2.3 **Technical Standardization of Artificial Intelligence**

(Raffaella D'Alessandro)



Impacts of Adversarial Use of Generative AI on Homeland Security

Preparedness Series
January 2025

Science and Technology

Security Critical Infrastructure: Real-world AI-synthesized Fake Media Scenario

A false report of an explosion at the Pentagon, accompanied by an AI-generated image, as illustrated in Figure 4.4, spread on Twitter in early 2023, causing a brief dip in the stock market before being debunked. The image, which depicted a black cloud of smoke near a building, was quickly identified by experts as likely being generated by AI. The Department of Defense and Arlington County Fire Department confirmed that there was no explosion or incident at the Pentagon.

Figure 3.4. A False Report of Explosion at Pentagon[140]

This incident highlights the growing threat of AI-generated deepfakes in spreading disinformation and manipulating public perception. The fake image was circulated by several Twitter accounts (now known as X), some of which had blue checkmarks — previously a symbol of verified accounts, but later obtainable by any user who subscribes to Twitter Blue. The hoax briefly impacted financial markets and demonstrated how easily deepfakes can be created and disseminated, raising concerns about the potential for future misuse of this technology.[141]

Guidance

National AI Strategy

The National AI Strategy builds on the UK's strengths but also represents the start of a step-change for AI in the UK, recognising the power of AI to increase resilience, productivity, growth and innovation across the private and public sectors.

21 MARZO 2025 - *London Heathrow Airport*

Substations, operated by **National Grid**, are designed to produce, convert, and distribute electricity at suitable voltage levels. **Heathrow uses three electricity substations, each with a backup.**

There are also **backup diesel generators**, and **uninterruptible battery-powered supplies** which provide enough power to keep safety critical systems such as aircraft landing systems running.

However, when the fire broke out the substation, it was out of action, along with its backup.

Heathrow's main fall-back was the **two remaining substations**, but the airport's CEO, Thomas Woldbye, told the BBC that it **"takes time" to "switch them".**

He said the incident was "not created at Heathrow Airport, it was created outside the airport and we had to deal with the consequences".

https://www.bbc.com/news/articles/cgm1krkxrxgo

More than **63,000** homes lost power

150 people were evacuated from surrounding properties.

Nearly **1,400 flights** were disrupted by the airport's closure

Around **120 flights** were diverted elsewhere.
Around **200,000 pax** stranded

84 mln pax (2024)

1300 dep/arr 24h

FCO Statistiche (2024)	
Passeggeri in transito	49.203.734 13,0% (vs 2019) - 1
Movimenti	315.597 1,9% (vs 2019)

The head of Heathrow Airport described Friday's closure of the airport as **"an incident of major severity"** as he apologized to passengers for the disruption.

"I like to stress that this is has been an incident of major severity. It's not a small fire. We have lost power equal to that of a mid-sized city, and our backup systems have been working as they should, but they are not sized to run the entire airport," CEO Thomas Woldbye said in a statement Friday.

"We will prioritize **incoming flights that have been stranded in Europe** so we can get as many passengers into the country as we possibly can. After that, a few flights will be taking off during the evening, and that is about it," Woldbye added.

“I’d actually warned Heathrow of concerns that we had with regard to the substations,” Nigel Wicking, the chief executive of the Heathrow Airline Operators’ Committee, which represents dozens of airlines that use the airport, told the parliamentary transport committee.

Mr. Wicking said that on March 15, less than a week before the major power loss at the airport, he had raised the issue of ***thefts of wires and cables*** *near an airport power supply that had briefly affected* ***some runway lights***.

“The incident Mr. Wicking referred to,” said Abbie Uzzell, “did not involve the three main incoming power supplies to the airport**. This issue related to a minor substation, of which there are 250 at the airport**. We were well aware of that incident before Mr. Wicking raised it.”

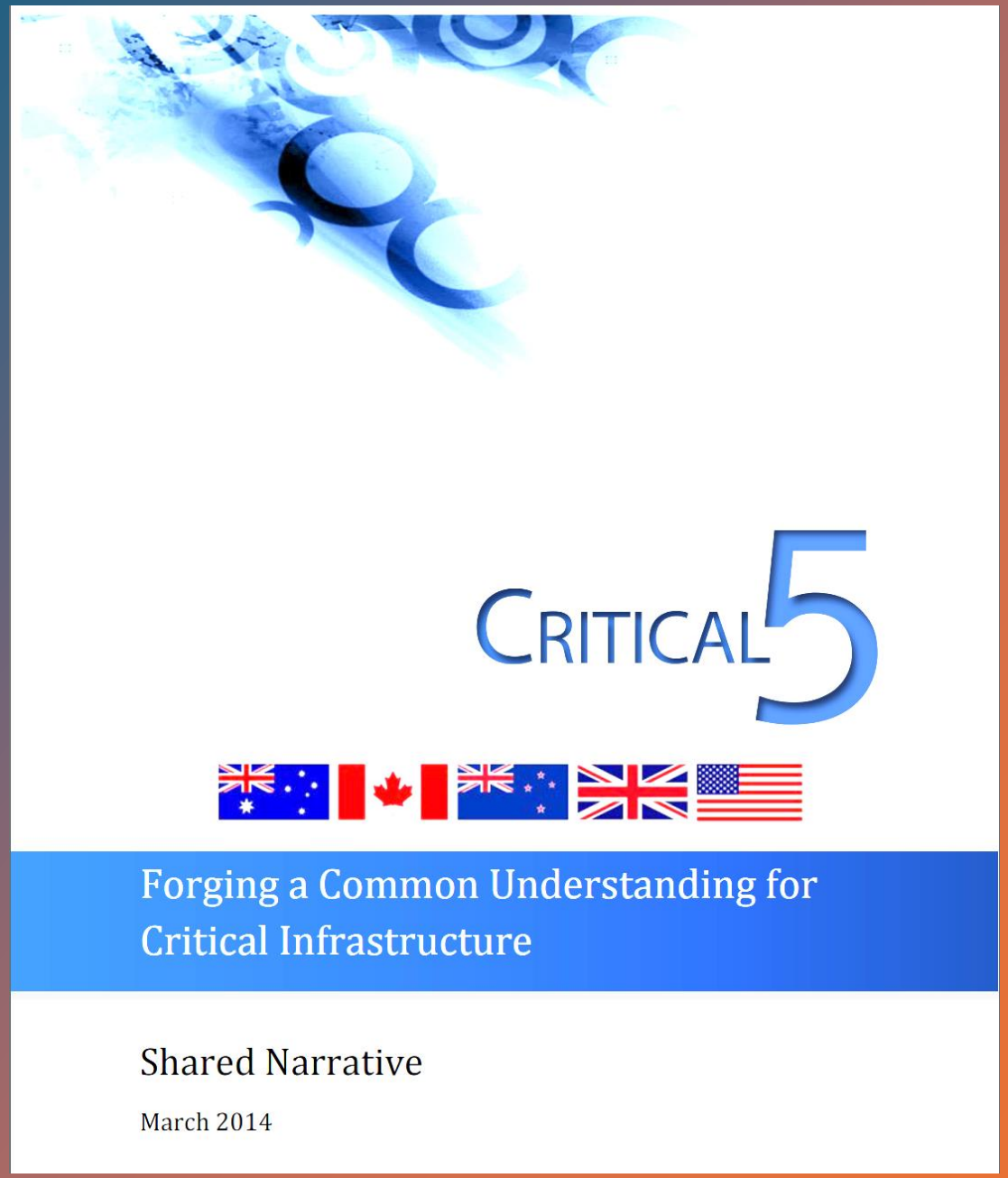
https://www.nytimes.com/2025/04/02/world/europe/heathrow-airport-power-outage-substation.html#

Distanza tra sottostazione e recinzione Apt: 2.464,92 m

https://www.bbc.com/news/live/cly24zvvwxlt

Investigation into power outage: Who's saying what?
published at 18:56 22 March
As we've been reporting, the government has ordered the **National Energy System Operator (NESO)** to conduct an investigation into the power outage at Heathrow Airport.
Let's recap what the key players have been saying:
Energy Secretary Ed Miliband says he will work in partnership with Ofgem* to "properly understand what happened and ***what lessons need to be learned***".
Transport Secretary Heidi Alexander says it is "imperative" to "identify how this power failure happened and learn from this ***to ensure a vital piece*** of national infrastructure remains strong".
Akshay Kaul, director general for infrastructure at Ofgem, says the regulatory authority will ***"not hesitate to take action***", adding that they will work to ensure NESO's review "goes as far as possible" to ensure steps are put in place to avoid similar events in the future.
Fintan Slye, Chief Executive of NESO - who have been commissioned to lead the investigation -welcomed the commission adding that the operator will work with ***"all relevant stakeholders to understand the lessons*** that can be learned to improve the future resilience of Britain's energy system".
Heathrow Airport's Chief Executive Officer Thomas Woldbye welcomed the investigation and hopes the findings will "***help strengthen the airport's future resilience"***.

ofgem * Office of Gas and Electricity Markets



Guidance
Critical 5 Joint Statement on Security and Resilience
Published 21 December 2024

- Building on the shared commitment to critical infrastructure security and resilience outlined in the recently released Adapting to Evolving Threats: A Summary of Critical 5 Approaches to Critical Infrastructure Security and Resilience, the Critical 5 reaffirms its vision of fostering collaboration across the private and government critical infrastructure communities in our five nations. Owners and operators will be better positioned to enhance their resilience by exchanging knowledge, experience, and expertise across a range of resilience measures, including:

- **Identifying interdependencies and priorities within critical infrastructure systems and assets;**
- Assessing current and emerging risks to understand the full spectrum of threats and hazards;
- Developing actionable incident response and recovery plans to rapidly restore operations;
- Evaluating and adapting those plans based on lessons learned from exercises and real-world events;
- Working proactively with each other, civil society, and all levels of government to ensure that critical infrastructure security and resilience is a whole-of-community effort.

UK AI policy

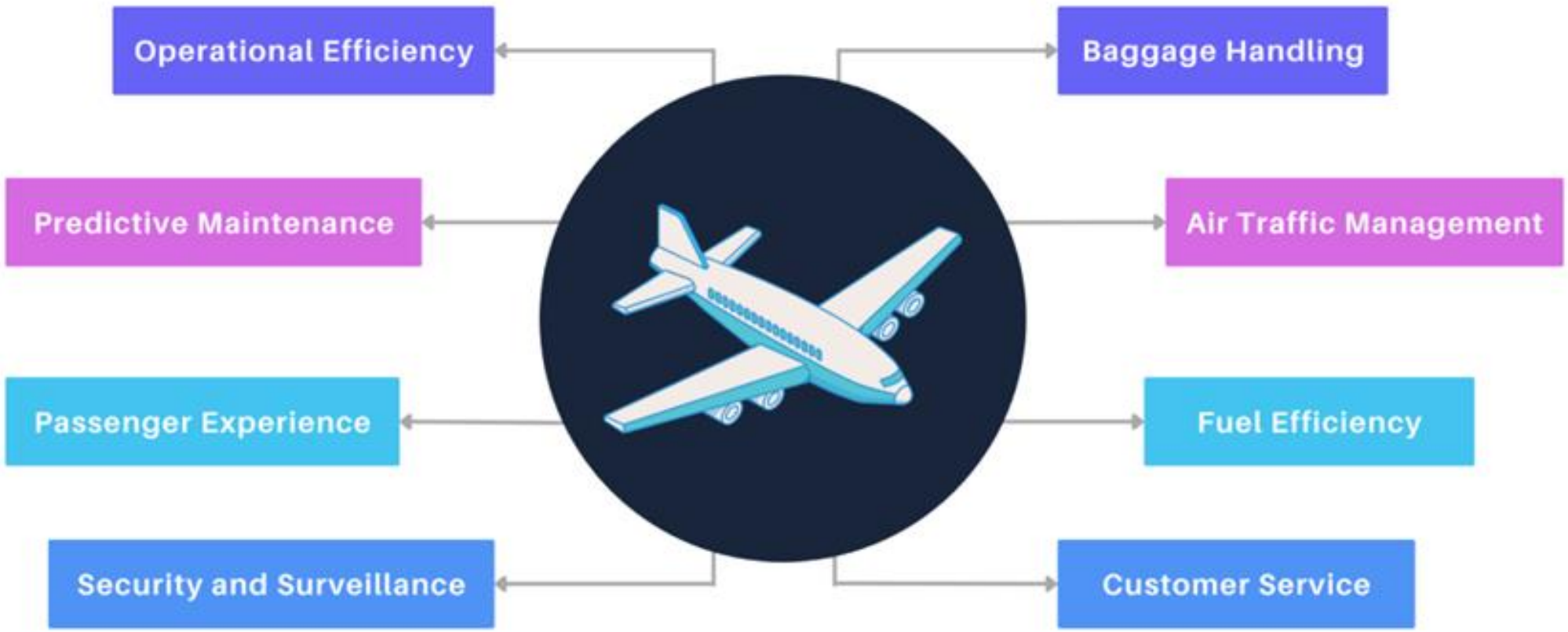
National AI Strategy

- The National AI Strategy builds on the UK's strengths but also represents the start of a step-change for AI in the UK, recognising **the power of AI to increase resilience**, productivity, growth and innovation across the private and public sectors.

- Long term (next 12 months and beyond):

Work with national security, defence, and leading researchers to understand what public sector actions can safely **advance AI and mitigate catastrophic risks**

ARTIFICIAL INTELLIGENCE IN AVIATION



AIIC-ISACAROMA 6 maggio 2025 "CRITICAL INFRASTRUCTURE RESILIENCE and ARTIFICIAL INTELLIGENCE"

Vincoli Normativi ed Etici non bastano ad assicurare la Resilienza.

il caso Heathrow e un auspicabile contributo della intelligenza artificiale

alberto.carusodecarolis@fmsconsulting.it
Mob. + 39 335 7206804