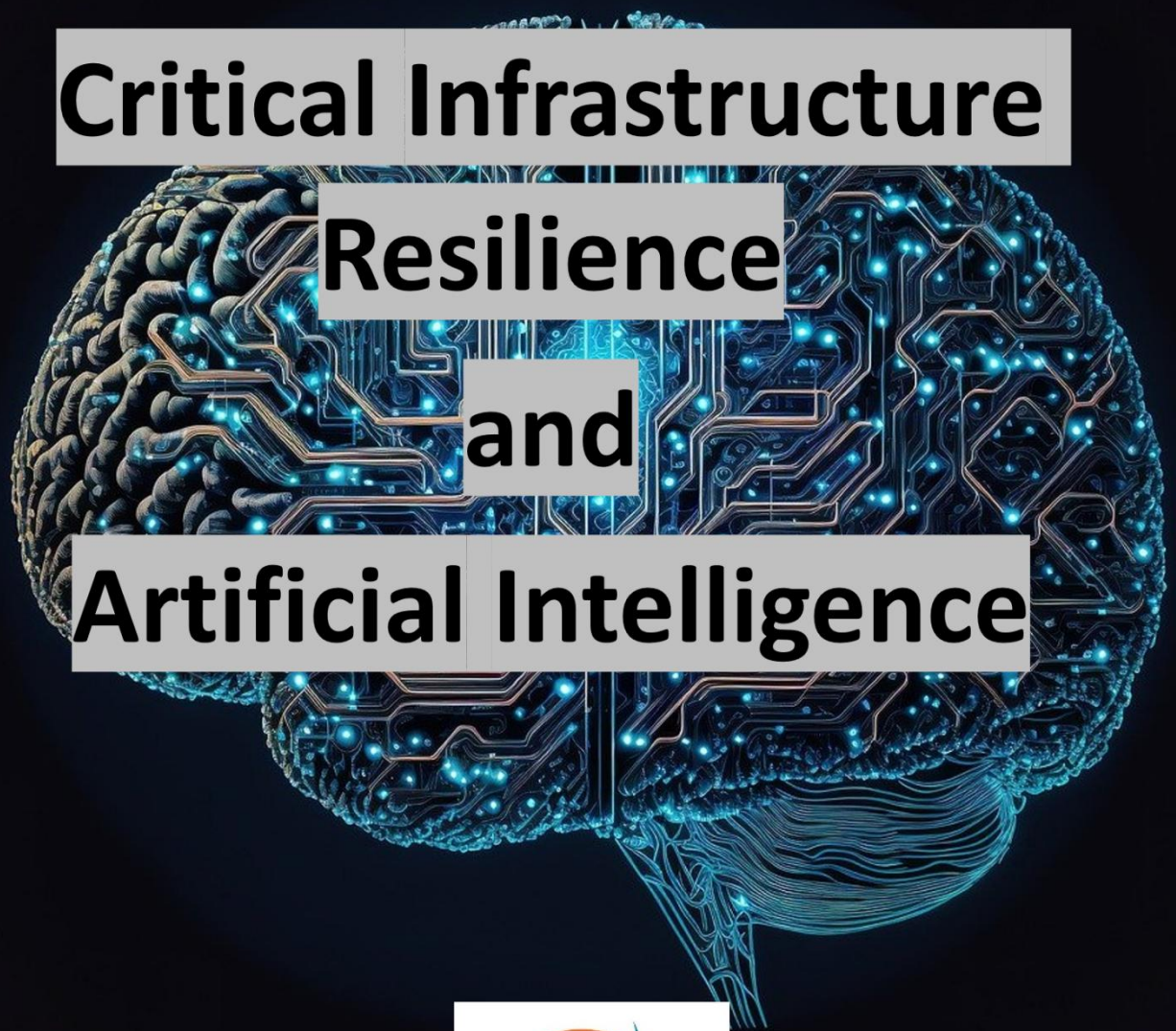
Critical Infrastructure Resilience and Artificial Intelligence



AIIC
ASSOCIAZIONE ITALIANA
ESPERTI IN INFRASTRUTTURE CRITICHE

Critical Infrastructure Resilience and Artificial Intelligence

Published by AIIC

February 2025

This document is the result of a joint project coordinated by Sandro Bologna and carried out with the contribution of *(in alphabetical order)* 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 Versaci.

The authors would like to thank the Italian Association of Critical Infrastructure Experts (AIIC) for its support and encouragement.

This version of the report represents the state of the art at the date of publication.

SUMMARY

CRITICAL INFRASTRUCTURE RESILIENCE AND ARTIFICIAL INTELLIGENCE

EXECUTIVE SUMMARY

This document 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 underscores the importance of a holistic and collaborative approach to leveraging AI's potential in critical infrastructure while addressing ethical, technical, and regulatory challenges.

The document adopts the 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, entered into force on 16 January 2023 and repealing Council Directive 2008/114/EC. The Directive aims to strengthen the resilience of critical entities against a range of threats, including natural hazards, terrorist attacks, insider threats or sabotage, and public health emergencies.

The document examines current policies and regulations governing AI use. It identifies regulatory gaps and proposes improvements to balance innovation and safety. Globally, nations and international organizations are developing strategies and standards to ensure safe and effective adoption of AI. Chapter 2 compares key initiatives, highlighting the importance of international collaboration in addressing shared challenges.

Various AI tools and algorithms have been used in the industrial world for many years in different application domains, but in a different way from today when the main goal is transforming critical infrastructure by enhancing efficiency and resilience. However, its adoption also introduces new vulnerabilities, particularly in sectors such as energy, transportation, and communications. Chapter 3 provides an overview of the history of the application of AI concepts in the industrial world.

Governance frameworks and regulatory models are essential for the effective deployment of AI. Chapter 4 proposes systems for monitoring, managing, and responding to crises by integrating AI technologies with traditional infrastructure systems.

As mentioned before AI can both mitigate risks and introduce new ones. Chapter 5 explores innovative methods for identifying, assessing, and managing risks associated with AI adoption in critical infrastructure.

Chapter 6 provides a critical perspective on existing AI applications, highlighting successes, limitations, and future opportunities in areas such as predictive maintenance and cybersecurity.

The introduction of AI raises ethical and societal issues, including transparency, accountability, and the impact on employment. Chapter 7 delves into human rights considerations and the need for ethical approaches in AI implementation.

Real-world case studies demonstrate how AI has been successfully (or unsuccessfully) implemented in critical infrastructure. Each example highlights lessons learned and best practices. These are reported in Chapter 8.

Appendix I includes technical details about the AI applications for resilience for each of the different categories of critical infrastructure as identified in the Directive (EU) 2022/2557[1].

[1] Directive on the Resilience of Critical Entities, https://eur-lex.europa.eu/eli/dir/2022/2557/oj

1 APPLYING ARTIFICIAL INTELLIGENCE TO CRITICAL INFRASTRUCTURE *(Sandro Bologna, Glauco Bertocchi)*

Artificial Intelligence (AI) technologies are increasingly being integrated to strengthen critical infrastructure (CI) against potential disruptions. Examples include: machine learning, predictive analytics, advanced AI algorithms for threat detection, risk assessment and adaptive response mechanisms. However, while AI is growing in popularity, it may not be able to replace human expertise and judgement in the near or even distant future, and it has vulnerabilities that can be exploited by cybercriminals. 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.

1.1 AI for what?

In the United States the Cybersecurity and Infrastructure Security Agency (CISA), has identified the ten most relevant usages of AI in CI[2]. This report is focused on using AI to improve CI's resilience and for a better focus, we have further grouped the uses into three macro categories, namely

1. **Predictive Maintenance**: AI can analyze vast amounts of data from sensors embedded in critical infrastructure systems to predict potential failures before they occur. By identifying issues early, maintenance can be performed proactively, reducing downtime and preventing catastrophic failures.
2. **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. This information can be used to prioritize investments in resilience measures and develop contingency plans.
3. **Security and Safety, specifically 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. This includes identifying suspicious network activity, analyzing malware, and automatically patching vulnerabilities.

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

1.2 The importance of data

Applying AI models in critical systems necessitates addressing several challenges, especially concerning the data used in these models (stochastic computations, machine learning, etc). Here are key aspects that need to be addressed:

1. **Data Quality**:
    - **Accuracy**: Ensuring the data is correct and represents the real-world scenario accurately.
    - **Completeness**: Making sure that the data is comprehensive and includes all necessary variables.

[2] https://www.dhs.gov/sites/default/files/2024-04/24_0426_dhs_ai-ci-safety-security-guidelines-508c.pdf

- **Consistency**: Ensuring that the data does not contain contradictions and maintains uniformity across different datasets.
- **Timeliness**: Using the most up-to-date data available to reflect current conditions accurately.

2. **Data Quantity**:
   - **Sufficiency**: Having enough data to train robust models. Insufficient data can lead to overfitting or underfitting.
   - **Diversity**: Ensuring the data covers a wide range of scenarios and edge cases to make the model generalizable.

3. **Bias and Fairness**:
   - **Bias Detection and Mitigation**: Identifying and reducing biases in the data to prevent discriminatory outcomes.
   - **Fairness**: Ensuring that the model's decisions are equitable and do not disproportionately affect any group.

Addressing these aspects ensures that AI models used in critical systems are reliable, fair, secure, and effective.

1.3 Relationship between Internet of Things (IoT), Big Data Analytics, Critical Infrastructure Resilience, and Artificial Intelligence (AI)

The relationship between the **Internet of Things (IoT)**, **Big Data Analytics**, **Critical Infrastructure Resilience**, and **Artificial Intelligence (AI)** is complex and synergistic, as these elements combine to enhance the management, security, and robustness of critical infrastructure while optimizing decision-making and operations. Here's an analysis of each concept and how they interconnect.

1.3.1 Internet of Things (IoT)

IoT represents the network of physical devices (sensors, actuators, smart devices, machinery) connected to each other and the Internet, which collect, process, and exchange data. In critical infrastructure (such as energy, transportation, healthcare, telecommunications, etc.), IoT enables real-time monitoring of various parameters (for example, power flow in an electrical grid or machine conditions in a factory). AIIC published in June 2020 the Report **"Internet of Things (IoT) in the context of Critical Infrastructure – an Enterprise Security and Management Perspective for a Security and Privacy Compliance",** analyzing IoT aspects and suggesting how to use them.

1.3.2 Big Data Analytics

IoT generates a large volume of real-time data from critical infrastructure. Big Data Analytics enables the analysis, extraction, and interpretation of these massive data volumes. Through advanced analytics, it's possible to identify patterns, predict faults, and optimize operations. For instance, in power grids, data collected from sensors can be used to identify malfunctions before service interruptions occur. AIIC published in February 2018 the Report **"Good Practices and Recommendations on the use of Big Data Analytics for Critical Infrastructure Resilience"**, analyzing Big Data Analytics aspects and suggesting how to use them.

1.3.3 Resilience of Critical Infrastructure

Critical infrastructure (like power grids, water systems, transportation, healthcare systems, telecommunication networks) is essential for the functioning of society and the economy. AIIC published in February 2016 the Report "**Guidelines for Critical Infrastructure Resilience Evaluation**" exploring the different dimensions of Resilience and its differences from the concepts of reliability mostly used at that time. Resilience is the ability to anticipate, absorb, adapt to, recover and operate even in the presence of disturbances or attacks. The combined use of IoT, Big Data Analytics, and AI significantly improves **resilience** by enabling for example:

- **Real-time monitoring** of infrastructure conditions.
- **Rapid response** to faults or anomalies through automated decision processes.
- **Prediction of critical events** (e.g., overloads, mechanical breakdowns, cyber-attacks) with predictive analytics.
- **Optimization of recovery processes** and resource management during and after an emergency.

At the beginning of the year 2024, AIIC decided to explore the contribution of AI to Critical Infrastructure Resilience.

1.3.4 Importance of Data

Data is foundational for deploying AI models in critical systems. When using AI for critical systems like healthcare, energy, transportation, telecommunications, or in general infrastructure management, careful attention to data quality, quantity, and ethical concerns like bias is essential to ensure reliability and trustworthiness. In the following of the present Report, we mention an expanded view on each aspect. Here we only wish to stress the concepts of **Bias and Fairness.** Ensuring ethical integrity and fairness is essential in AI, especially in critical systems impacting diverse populations:

- **Bias Detection and Mitigation**: Bias in data can lead to discriminatory outcomes, as AI models trained on biased data may inadvertently reinforce those biases. Identifying biases, whether they're due to historical inequities or data collection practices, and actively mitigating them are essential steps.
- **Fairness**: Fairness ensures that model outcomes are equitable across all user demographics, minimizing any negative impacts on vulnerable populations. In healthcare, for example, a fair model would ensure that predictions and treatment recommendations are accurate and accessible across different patient groups without unintended discrimination.

1.3.5 Synergy Between IoT, Big Data, and AI for Improved Resilience

The combined use of IoT, Big Data Analytics, and AI results in a synergistic approach to infrastructure resilience:

- **Enhanced Situational Awareness**: IoT continuously monitors infrastructure conditions, while Big Data Analytics processes this data, and AI identifies patterns and risks. This combination provides operators with a clearer understanding of real-time infrastructure health and the ability to anticipate and manage potential threats.
- **Proactive Decision-Making**: By leveraging AI-driven predictive analytics, infrastructure managers can identify high-risk components, optimize maintenance schedules, and allocate resources efficiently, which significantly reduces the likelihood of catastrophic failures.
- **Automated Recovery and Adaptation**: In the event of an infrastructure failure, AI algorithms can automatically take corrective actions. For example, in a smart power grid, AI

might reroute power to unaffected areas, isolate damaged sections, and prioritize resources for rapid recovery.

1.3.6 Challenges and Considerations

Despite its promise, the integration of IoT, Big Data, and AI for critical infrastructure resilience faces several challenges:

- **Data Privacy and Security**: IoT networks are often vulnerable to cyber-attacks, and the data they generate may include sensitive information. Ensuring data protection and secure communication is critical to prevent data breaches or malicious manipulation.
- **Interoperability**: Different IoT devices and analytics systems often use varied standards, which can complicate integration. Standardization and interoperability are essential to ensure that devices and systems can work seamlessly together.
- **Reliability and Latency**: Real-time decision-making requires reliable, low-latency communication. Infrastructure resilience systems must have robust, high-speed networks to ensure that IoT data reaches analytics platforms and AI systems with minimal delays.
- **Scalability and Cost**: Implementing IoT, Big Data, and AI across large infrastructure networks can be costly. Scalability is a challenge, especially for public infrastructure agencies with limited budgets. Modular approaches and cost-efficient technologies are needed for widespread adoption.

In summary, the integration of IoT, Big Data Analytics, and AI marks a significant advancement in critical infrastructure resilience, enabling predictive, preventive, and autonomous capabilities that allow infrastructure to withstand and quickly recover from disruptions, once we can demonstrate that we have faced and solved (or at least reduced to an acceptable level) the above listed problems.

1.3.7 Use Cases of Integrated IoT, Big Data Analytics, and AI in Critical Infrastructure

Several real-world applications illustrate how the integration of these technologies improves resilience:

- **Smart Power Grids**: IoT sensors in the grid monitor electricity flow, demand, and equipment health. Big Data Analytics processes this data to detect patterns and predict failures. AI systems then automatically adjust power distribution, balance loads, and manage outages to minimize downtime.
- **Transportation Systems**: IoT enables real-time monitoring of bridges, roads, and public transit systems. Analytics systems can identify congestion patterns and predict maintenance needs. AI can optimize traffic flow, reroute vehicles, and coordinate emergency response, minimizing disruptions.
- **Healthcare Facilities**: In hospitals, IoT devices monitor equipment and environmental conditions. Big Data Analytics and AI can predict when critical equipment, such as ventilators, will need maintenance and ensure their availability during emergencies. AI algorithms can also assist in resource allocation, ensuring beds, supplies, and staff are optimally distributed.
- **Smart Water Networks**: In recent years, research in the sector has made enormous strides: artificial intelligence techniques, together with the implementation of intelligent sensors, can offer a significant contribution to the automatic monitoring of water losses, helping to make water networks more resilient and sustainable and reducing waste. Furthermore, machine learning models, based on the huge amount of data available from *smart meters*, i.e. intelligent

meters that allow precise measurements at a distance, allow overcoming traditional methods of finding leaks, such as visual inspections, analysis of acoustic signals and vibrations[3].

Examples of these four real-world applications have been illustrated in chapter 8 of this Report.

[3] https://serviziarete.it/wp-content/uploads/2023/09/luglio-agosto-2023-Lintelligenza-artificiale-per-le-reti-idriche.pdf

2 POLICY AND REGULATIONS: WORLDWIDE OVERVIEW STRATEGIES ON IA & STANDARDS. *(Alberto Caruso de Carolis, Raffaella D'Alessandro, Luisa Franchina, Adriana Peduto, Maria Beatrice Versaci)*

This chapter illustrates the national, European Union and NATO visions, on strategies for the development and use of artificial intelligence, through a brief review of international literature and official documents.

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)*

The growing adoption of artificial intelligence (AI) promises to enhance the efficiency, resilience, and security of critical infrastructure but also introduces new risks and challenges. To address these challenges, it is essential to establish a solid regulatory framework governing both AI implementation and the overall resilience of critical infrastructure.

A shared system of standards and regulations offers multiple benefits: fostering international cooperation, ensuring compliance with industry best practices, and strengthening stakeholder trust. Moreover, a well-defined regulatory framework mitigates risks associated with technological vulnerabilities, data misuse, and cyberattacks, ensuring AI operates transparently and ethically. Conversely, inadequate regulations could lead to uneven adoption of AI technologies, creating security gaps and systemic inefficiencies.

In today's technological and geopolitical landscape, the increasing digitalization and interconnection of critical infrastructure expose them to a growing range of threats, including cyberattacks, extreme natural events, and other systemic crises. Resilience, defined as the ability of infrastructure to withstand, absorb, and recover from adverse events, has thus become a strategic priority. Achieving this objective requires the establishment of adequate standards and regulations, providing a solid and shared framework with clear guidelines to ensure security, reliability, and operational efficiency. These tools enable organizations to adopt risk management practices aligned with established security models, supporting operational continuity and the ability to respond swiftly and effectively to crises and attacks.

The current regulatory framework, at both national and international levels, is rapidly evolving to keep pace with technological advancements, particularly to regulate the adoption and use of artificial intelligence (AI) in critical infrastructure.

In Europe, the Artificial Intelligence Act (AI Act) represents the first comprehensive attempt to regulate AI development and use, adopting a risk-based approach. Its strength lies in a tiered system classifying AI applications by their potential impact, enabling proportional regulation. However, its weaknesses include implementation complexity and potential rigidity, which might hinder innovation, particularly in the early stages of technological development.

Focusing specifically on AI applications within critical infrastructure, ENISA (European Union Agency for Cybersecurity) has issued specific guidelines for the secure integration of AI, recognizing its enormous potential to enhance operational efficiency and predictive capabilities, while also highlighting the risks of manipulation or attacks. The report *"Cybersecurity of AI and Standardisation"* emphasizes the importance of developing robust security standards to mitigate emerging risks associated with increasing AI adoption.

Additionally, international ISO standards for security and resilience, such as the ISO/IEC 42001:2023 standard for AI management systems, provide a shared framework for operational continuity, information security, and risk management. These standards are crucial for promoting a coordinated global response, as threats to critical infrastructure transcend national borders and require international cooperation for effective mitigation.

Italy's *National Artificial Intelligence Strategy (2020)* and AGID guidelines provide a reference framework for AI development at the national level, focusing on ethical and secure adoption of technologies. Key strengths include an emphasis on sustainability and public-private collaboration. However, potential weaknesses lie in the practical implementation of these directives, which may slow the country's competitiveness compared to other nations.

Looking beyond Europe, the U.S. has introduced the *Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government (2020)* and the *Algorithmic Accountability Act*, which aims to promote the responsible use of AI in the public sector and ensure algorithmic transparency. These legislative tools underscore the importance of accountability and trust in automated systems, a significant strength for critical infrastructure. However, regulatory fragmentation across states and levels of government may hinder harmonization of practices.

Other regulations, such as the UK's *Data Protection Act (2018)* and the *UK AI Strategy*, offer a framework for protecting individual rights while fostering reliable AI technology development. Additionally, the EU's *Directive on Automated Decision-Making* seeks to regulate algorithmic decision-making transparently and responsibly. However, these regulations must balance protecting individual rights with fostering innovation, as excessive rigidity could stifle technological advancement.

Globally, the *OECD AI Principles (2019)* and the *UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)* highlight the importance of ethical standards and fairness in AI use, advocating for a values-based approach centered on human rights and transparency. While these principles are critical for promoting ethical AI adoption worldwide, practical implementation varies widely among countries, potentially creating disparities in regulatory effectiveness.

Finally, international technical standards, such as those established by ISO/IEC JTC 1/SC 42, provide a critical reference point for the secure and interoperable development of AI technologies. These standards serve as a robust foundation for harmonizing global practices, though the primary challenge remains in ensuring their adoption and compliance by all nations.
However, these proposals fail - or address only at an embryonic stage - the issue of AI's specific impact on critical infrastructure. A dedicated regulatory intervention should explicitly address at least the following points:

1. **Creation of an adaptive, risk-based regulatory framework:** Inspired by the EU AI Act, such a framework should be tailored to critical infrastructure. It should define various risk classes for AI applications in sectors like energy, transportation, telecommunications, and healthcare, with compliance requirements varying according to risk levels. A clear

classification system, assessing AI's impact on these sectors, would enable continuous monitoring and greater flexibility in adapting regulations to technological advancements and emerging threats.

2. **Development of specific security and resilience standards for AI in critical infrastructure:** These standards should mandate transparency and algorithm verifiability, ensuring AI models are not only effective but also robust against cyberattacks and manipulations. Regulations must establish parameters to ensure AI systems can operate under stress or emergency conditions, maintaining continuity of essential services.

3. **Integration of AI-specific cybersecurity regulations:** The rapid proliferation of AI in critical infrastructure introduces new attack surfaces requiring targeted cybersecurity regulations. Current frameworks should be expanded to address unique AI-related threats, such as attacks on machine learning algorithms or training data manipulation. New regulations should mandate advanced defense mechanisms from AI solution providers.

4. **Promotion of interoperability and global standards:** A growing challenge in AI adoption for critical infrastructure is regulatory fragmentation across countries and industries. To address this, a regulatory framework promoting interoperability among AI systems globally should be adopted, supported by international standards. The goal is to ensure critical infrastructure, interconnected globally, can be managed securely and consistently worldwide.

2.2 Impact of Regulations on AI Adoption *(Alberto Caruso de Carolis, Adriana Peduto)*

The introduction of new regulations targeting artificial intelligence (AI) in critical infrastructure will have a significant impact on the adoption of these technologies. While the proposed regulations aim to ensure safety, fairness, and transparency, they are inevitably associated with additional costs and potential operational obstacles, but also long-term benefits for the resilience and trust in critical infrastructure.

2.2.1 Italy[4]

Italy's AI Strategy for 2024-2026: The Key Points[5]

The strategy highlights the importance of scientific research in improving the quality of life and the social environment. The actions proposed in this regard include the consolidation of an Italian AI research ecosystem that facilitates the exchange of knowledge between universities, research centers, and businesses. Such an ecosystem is also expected to be a breeding ground for the development of innovative start-ups, the support of a plan to retain and attract talent, the development of national AI Large Language Models that respect the values of European regulations, and the funding of blue-sky research for next-generation AI.

AI is also seen as a critical tool in the transformation of the public administration to improve internal efficiency and provide services tailored to citizens' needs. To fully exploit AI's potential, a structured and systematic approach becomes necessary, including actions to guarantee privacy, security, and proper data management, as well as the development of AI systems for interoperability and training

[4] https://oecd.ai/en/dashboards/countries/Italy

[5] Giacomo Lusardi and Alessandra Faranda, 14 April 2024, https://blogs.dlapiper.com/iptitaly/2024/04/italys-ai-strategy-for-2024-2026-the-key-points/

of public personnel. In addition, guidelines should be adopted to promote the use of AI in public tenders and create AI applications for the public sector that can guarantee adherence to regulations. According to the executive summary, AI could also simplify the interaction between public authorities and citizens or businesses by developing large-scale solutions based on feedback and specific needs. Finally, the strategy urges comprehensive training on AI in public administration through upskilling courses for staff.

Regarding businesses, the strategy aims to shed light on AI's benefits to the Italian production and entrepreneurial system, known for its process and product excellence, and manufacturing vocation. A twofold strategic approach is proposed: on the one hand, the role of Italian ICT companies in the development of AI systems should be enhanced by fostering collaboration with universities and research bodies and facilitating the management of regulatory and certification practices; on the other hand, companies not directly involved in technological development but influenced by AI should align their strategies towards a greater centrality of data and AI to increase their competitiveness, with a particular focus on the challenges of environmental sustainability. The strategy proposes coordinated actions to strengthen the AI ecosystem among SMEs through dedicated funding to support the adoption and development of interoperable AI solutions. It also highlights the need to create laboratories to develop AI applications in industrial contexts and to support the growth of start-ups operating in the sector.

Last but not least, AI training. The executive summary notes that there is currently a shortage of AI skills in Italy, which slows the adoption of innovative solutions. To this end, the strategy proposes an integrated plan to strengthen and spread knowledge of AI in the education system, from high schools to universities, paying particular attention to PhD programs. Furthermore, structured reskilling and upskilling programs in both the public and private sectors are envisaged to update skills and retrain workers to use new technologies. Similarly, promoting AI literacy for the population becomes essential to avoid creating a knowledge gap that undermines social and economic cohesion in the long run. In this respect, the strategy proposes implementing AI learning paths in schools, creating internships, exchange and visiting programs in companies and research centers, introducing AI as a subject in university degree courses, and supporting the National PhD in AI.

2.2.2 European Union[6]

How the EU Can Navigate the Geopolitics of AI[7]

The surge in AI innovation has prompted a parallel race in crafting regulatory frameworks. The absence of a comprehensive global governance structure has led to a proliferation of international, European, and national initiatives, non-binding principles and norms, and voluntary corporate codes of conduct, forming a complex regulatory and governance landscape. The EU's landmark AI Act aims to set a precedent for a binding hard regulation of AI, reflecting a commitment to human-centric, trustworthy, and risk-based regulation.
While the EU has had a head start and first-mover advantage in setting the global agenda with the AI Act as a blueprint for other governments, the current race to govern and regulate AI highlights an increasingly crowded AI governance landscape.
This will be difficult for the EU to navigate, as it cannot solely rely on the "Brussels effect" and extraterritorial regulations to influence international standards. The union's unique governance model and emphasis on democratic values may clash with the diverse regulatory and innovation approaches of other regions. Further challenges facing the EU include coordinating actions across its institutions

[6] https://oecd.ai/en/dashboards/countries/EuropeanUnion
[7] Raluca Csernatoni, published on January 30, 2024, https://carnegieendowment.org/europe/strategic-europe/2024/01/how-the-eu-can-navigate-the-geopolitics-of-ai?lang=en

and member states while crafting a cohesive European AI foreign policy approach, both in terms of countering U.S.-driven innovation dominance and China's quest to become an AI superpower. Not to mention growing AI nationalism, protectionist tendencies, and fragmentation risks within the bloc.

To address the AI innovation lag in Europe, on January 24, 2024, the European Commission unveiled a comprehensive AI innovation package—an important move toward fostering a dynamic and robust AI ecosystem in Europe. To further boost the leadership of European start-ups and cultivate competitive AI ecosystems across the union, the Commission plans to establish what it terms "AI Factories," comprising AI-dedicated supercomputers, interconnected data centers, and a skilled workforce ranging from supercomputing and AI experts to data specialists, researchers, and start-ups.

These measures, in the wake of the political consensus achieved in December 2023 on the AI Act, are explicitly designed to propel the creation, implementation, and adoption of trustworthy AI within the EU. Yet, the proof is in the proverbial pudding when it comes to such initiatives. The crucial test lies in translating these commitments into tangible actions, particularly in terms of fostering a vibrant and globally competitive cross-border AI innovation (start-up) ecosystem in the EU.

Establishing an AI Office within the Commission could help ensure a more streamlined development and coordination of AI policy at the European level, as well as supervise the implementation and enforcement of the AI Act.

The EU faces the challenge of managing the geopolitics of AI governance in a landscape characterized by state and corporate competition, as well as an emerging global regime of complex regulatory frameworks. While the EU's commitment to responsible AI is commendable, building a harmonized European AI foreign policy approach, fostering strategic alliances with key partners, effectively operationalizing the AI Act, and navigating diverse governance initiatives will be crucial for shaping the future of AI on a global scale.

2.2.3 NATO[8]

NATO releases revised AI strategy[9]

On Wednesday (10 July 2024), NATO released its revised artificial intelligence (AI) strategy, which aims to accelerate the use of AI technologies within NATO in a safe and responsible way.
It builds on one published in 2021 and takes account of recent advances in AI technologies, such as generative AI, and AI-enabled information tools.

The strategy identifies several priorities, including: advancing the implementation of NATO's Principles of Responsible Use; increasing interoperability between AI systems throughout the Alliance; the combination of AI with other emerging disruptive technologies; and expanding NATO's AI ecosystem through closer cooperation with Allied industry and academia, NATO's Defence Innovation Accelerator DIANA, the NATO Innovation Fund and like-minded partners.

For the first time, the strategy also identifies AI-enabled disinformation, information operations and gender-based violence as issues of concern for the Alliance, our societies and democracies.
Under the new AI strategy, NATO will work to protect against the adversarial use of AI, including through increased strategic foresight and analysis.

[8] https://www.nato.int/cps/en/natohq/official_texts_227237.htm.
[9] https://www.nato.int/cps/en/natohq/news_227234.htm.

2.2.4 USA

AI in the United States[10]

"The United States works with domestic and international AI communities to establish frameworks that advance trustworthy AI for all[11]."

Over the past year, dramatic advances in generative Artificial Intelligence (AI) and its rapid availability in products and services have catapulted AI into the public's imagination with images and predictions that are both enormously promising and deeply concerning. To respond to these trends, the United States has sought to address AI technologies holistically by focusing on the potential of AI to boost economic prosperity, help overcome major societal challenges and close the digital divide. It also recognizes that trustworthy AI requires strong governance tools that engage all relevant stakeholders to create a safe, secure and productive AI ecosystem.

As a starting point, in 2022, the White House published a Blueprint[12] for an AI Bill of Rights to help counter harms that AI can perpetuate, including discrimination in hiring processes or credit decisions and violations of individual privacy. The Blueprint identified five principles to guide the use and development of AI, in large part inspired by the OECD AI Principles[13] that the United States has endorsed. In January 2023, the U.S. Department of Commerce's National Institute of Standards and Technology (NIST) also released its AI Risk Management Framework (AI RMF)[14].

More recently, on October 30th, President Biden issued a landmark Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Artificial Intelligence to ensure that the United States leads the way in seizing the benefits and managing the risks of AI. The E.O. aims to establish new standards for AI safety and security, protect Americans' privacy, advance equity and civil rights, stand up for consumers and workers, promote innovation and competition, and advance American leadership around the world. Importantly, the E.O. also builds on previous White House work to establish voluntary commitments from 15 leading U.S. companies to drive safe, secure, and trustworthy development of AI.

Safeguarding Critical Infrastructure[15] (DHS)

AI-synthesized fake media challenges the security and integrity of critical infrastructure. One significant risk is the use of fake signals for water treatment plants. By generating false but realistic signals used by sensors, attackers can trigger unwarranted shutdowns, such as manipulating chemical readings to indicate dangerous levels, thereby disrupting essential water services. Similarly, AI-enhanced disinformation campaigns can manipulate the public into behaviors that endanger critical infrastructure. For instance, a deepfake government official issuing a fake evacuation notice could create massive traffic congestion, hindering emergency services and access to vital facilities.

[10] https://oecd.ai/en/dashboards/countries/UnitedStates
[11] Elham Tabassi, Isabel Gates, Sam Schofield, December 20, 2023, https://oecd.ai/en/wonk/united-states-ai-for-all-policy.
[12] https://www.whitehouse.gov/ostp/ai-bill-of-rights/.
[13] https://oecd.ai/en/ai-principles.
[14] https://www.nist.gov/itl/ai-risk-management-framework.
[15] https://www.dhs.gov/sites/default/files/2025-01/25_0110_st_impacts_of_adversarial_generative_aI_on_homeland_security_0.pdf

2.2.5 United Kingdom

AI in the United Kingdom[16]

'If it isn't diverse, it isn't ethical': the United Kingdom's approach to AI policy[17].

It isn't often that we hear such a complex issue in the social aspects of technology articulated so succinctly and with such clarity. In a world where governments and organisations rightly see the need to set out their own approaches to AI in ways that reflect their values and culture, establishing or signing up to principles as they do, the complexity of the issues they grapple with often makes implementing changes to make technology ethical very difficult. And yet, we can all agree, that a way to make artificial intelligence more suited to the society it serves, and more ethical, is to make sure it is being designed and built by diverse teams that reflect those societies.

The quote *'If it isn't diverse, it isn't ethical'* came from Professor Dame Wendy Hall, UK AI Council Skills Champion and author of the independent UK AI Review, published in October 2017.

The accelerated impact of AI and government action

Artificial intelligence is perhaps unique among technologies for its propensity to be self-reinforcing in both its development and use. This potential for accelerated impact makes ethical considerations all the more important, which is why the UK has set up various AI institutions in government and empowered regulators to consider how AI affects their own areas.

In the UK, the government has established several institutions tasked with thinking about AI, all of which were announced in 2017. The Office for AI[18] and AI Council[19] both came from recommendations[20] of the AI Review previously mentioned, with the former consisting of a team of civil servants tasked with taking forward AI policy and delivery, and the latter comprising an expert group of 22 leaders from AI industry, R&D and academia, and civil society. These institutions are complemented by the Centre for Data Ethics and Innovation[21], an expert advisory body that is independent of Government, and that draws on its own expert board and full-time staff. The Centre was not a recommendation of the AI Review but of the Royal Society's and British Academy's Data Governance report, also published 2017. In fact, ethics was not a focus of the AI Review, partly because the Centre for Data Ethics – which advises Government and Regulators on the use of data, including for (but not limited to) applications in AI – was conceived as the data stewardship body that would tackle all the ethical questions that AI raises.

2.2.6 Russian Federation[22]

Developing Artificial Intelligence in Russia: Objectives and Reality[23]

[16] https://oecd.ai/en/dashboards/countries/UnitedKingdom
[17] Stefan Janusz, Helen Embleton, June 1, 2021, https://oecd.ai/en/wonk/uk-ai-strategy
[18] https://www.gov.uk/government/organisations/office-for-artificial-intelligence
[19] https://www.gov.uk/government/groups/ai-council
[20] https://www.gov.uk/government/news/leading-experts-appointed-to-ai-council-to-supercharge-the-uks-artificial-intelligence-sector
[21] https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation
[22] https://oecd.ai/en/dashboards/countries/RussianFederation.
[23] Nikolai Markotkin and Elena Chernenko, August 5, 2020, Developing Artificial Intelligence in Russia: Objectives and Reality, (https://carnegieendowment.org/posts/2020/08/developing-artificial-intelligence-in-russia-objectives-and-reality?lang=en&center=russia-eurasia).

Even if AI development becomes Russia's highest priority, Moscow has no chance of catching up with Washington and Beijing in this field. Under favourable conditions, however, Russia is quite capable of becoming a serious player and even a local leader in certain areas.

The national strategy stresses two defining time markers for the development of AI in Russia: 2024, by which time Russia is expected to have significantly improved its positions in this field, and 2030, when it should have eliminated its lag behind developed countries and attained global leadership roles in certain AI-related areas. According to the document, Russia's key AI development priorities include increasing the number of entities involved in technological innovation by 50 percent, and creating a high-performance export-oriented sector equipped with modern technologies in key industries, primarily in manufacturing and agriculture.

The strategy focuses heavily on support for both state and privately sponsored scientific research. Citations of Russian scientific publications in the field of AI, as well as the number of patents and applied technological solutions registered and developed by Russian scientists, are expected to increase significantly by 2024. The government is also working to overhaul existing legal norms to simplify the development and implementation of AI-based technologies (for instance, driverless transportation). The document stresses the need for Russia's international cooperation on issues of AI-based product standardization and certification.

2.2.7 China[24]

Implications of China's AI Strategy: State Engineering, Domestic Challenges, and Global Competition[25]

China's artificial intelligence (AI) strategy represents a strategic blend of government-led initiatives and national development goals, aiming to establish a substantial presence in the global AI market. Characterized by extensive government investment, a domestically led tech ecosystem, and sector-wide AI integration, this strategy is rapidly advancing China's position as a technological superpower. Moreover, China's pursuit of AI leadership is reshaping China's technological and socioeconomic landscape, with significant implications for global power, global economic dynamics, and global governance of cutting-edge technologies.

2.3 Technical Standardization of Artificial Intelligence *(Raffaella D'Alessandro)*

The technical standardization of artificial intelligence is carried out at international level in ISO/IEC JTC 1 SC 42 (ISO/IEC JTC 1/SC 42 - Artificial intelligence) and at European level in CEN/CENELEC JTC 21 (https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/).

At Italian level these two committees are managed by UNI through UNI/CT 533 established at UNINFO, the Federated Body that has the delegation on the technical standardization of information technology (https://www.uninfo.it/).

[24] https://oecd.ai/en/dashboards/countries/China.

[25] Dr. Lizzi C. Lee, CCA Affiliated Researcher on Chinese Economy, February 21st, 2024, https://asiasociety.org/policy-institute/implications-chinas-ai-strategy-state-engineering-domestic-challenges-and-global-competition.

2.3.1 International Standardization of AI: ISO

In 2018, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) started a project on AI standardization by founding the subcommittee ISO/IEC JTC1 / SC42 Artificial intelligence. The Technical Management Board (TMB) of ISO decided that the Joint Technical Committee Information Technology" (JTC 1) should found a subcommittee (SC) on Artificial Intelligence. The inaugural plenary meeting of the new SC 42 took place in Beijing, China, in April 2018.

ISO/IEC JTC 1 SC 42 (SC42 for short) has a dual purpose: to produce the horizontal reference standards for AI and to be the point of reference for the other ISO, IEC and ISO/IEC JTC1 committees that produce vertical standards for AI. 38 permanent members and 24 observer members participate in the work of SC42 and the secretariat is entrusted to the United States. Italy has participated in the work as a "P" member since the establishment of SC42. The committee is structured in 5 WG (working groups), 4 JWG (Joint Working Groups) and 2 AhG (Ad Hoc Groups).

The WGs are: WG 1 – "Foundational Standards", WG 2 – "Big Data", WG 3 – "Trustworthiness", WG 4 – "Use cases & Applications" and WG 5 – "Computational approaches and characteristics". The JWGs are mixed working groups with: JWG2 – ISO/IEC JTC 1 SC 7 for "Testing of AI-based systems", JWG 3 – ISO/TC 215 for "AI enabled health informatics" JWG 4 – IEC TC65/65A for "Functional safety and AI Systems" JWG 5 – ISO TC 37 for "Natural language processing". The two ad hoc groups follow: AhG 4 – IT security issues together with ISO/IEC JTC 1 SC 27 AhG 7 – CEN/CLC JTC 21 projects for possible joint development.

This subcommittee has already produced 20 standards and has another 35 in production. Among those produced and published are: ISO/IEC 22989 "AI concepts and terminology", ISO/IEC 23894 "Guidance on risk management", ISO/IEC TR 24027 "Bias in AI systems and AI aided decision making", ISO/IEC TR 24028 "Overview of trustworthiness in artificial intelligence", ISO/IEC TR 20547 "Big Data Reference architecture", ISO/IEC 25059 "Quality model for AI systems". Many other standards are in progress or about to be published, you can follow the work program and publications at the following link: ISO/IEC JTC 1/SC 42 - Artificial intelligence.

2.3.2 European Standardization of AI: CEN/CENELEC

The AI Act (but also the Data Act and the Cyber Resilience Act) follow the principles of technical regulation in Europe with a risk-based approach: the essential requirements that the technology must meet are established by law (in this case the AI ACT); compliance with the requirements can be met with harmonized European standards. The "supplier" performs a conformity assessment based on the harmonized standards to demonstrate compliance. This reference model for product compliance in the EU is called the New Legislative Framework (NLF).

The AI Act provides for a series of harmonized European standards that make operational the common mandatory requirements applicable to the design and development of certain AI systems before they are placed on the market and that harmonize the methods of carrying out ex-post controls.

The "essential requirements" established in the AI Act are: Risk management system, Data and data governance, Technical documentation, Record-keeping, Transparency and provision of information to users, Human oversight, Accuracy, robustness and cybersecurity and have been "translated" into the request to develop harmonized standards relating to: Risk management system for AI systems, Governance and quality of datasets used to build AI systems, Record keeping through builtin logging capabilities in AI systems, Transparency and information to the users of AI systems, Human oversight

of AI systems, Accuracy specifications for systems, Robustness specifications of AI systems, Cybersecurity specifications of AI systems, Quality management system for providers of AI system, including post-market monitoring process, Conformity assessment for AI systems. The two European standardization bodies CEN and CENELEC have received the request to define these harmonized standards by April 2025 and will do so through the CEN/CENELEC JTC 21 “AI”.

JTC21, whose purpose is similar to that of ISO SC42, must adopt SC42 standards if they exist and must produce standards in line with European Union legislation, rules and principles. This committee is structured into 5 WGs (working groups): WG 1 – “Strategic Advisory Group”, WG 2 – “Operational Aspects”, WG 3 – “Engineering Aspects”, WG 4 – “Foundational & Societal Aspects”, WG 5 – “Cybersecurity. Among the standards that have been published by JTC1 we remember the ISO/IEC standards: ISO/IEC EN 22989 and ISO/IEC EN 23894.

Many other standards are in progress or about to be published, you can follow the work program and publications at the following link: https://standards.cencenelec.eu/dyn/www/f?p=CEN:84 .

3 THE ROLE OF ARTIFICIAL INTELLIGENCE IN CRITICAL INFRASTRUCTURE[26] *(Alberto Stefanini)*

3.1 Introduction

Artificial Intelligence (AI) has progressively transformed from an abstract notion into an essential component of modern critical infrastructure, including energy, transportation, telecommunications, and healthcare systems. Originally developed as a field exploring whether machines could mimic human thought processes, AI has since expanded to support complex applications in domains that require high levels of reliability, resilience, and adaptability. In these contexts, AI has become pivotal for optimizing performance, predicting and managing risks, and enhancing security across sectors critical to societal well-being.

The journey of AI began in 1950 with Alan Turing, a visionary in computer science, who posed a profound question in his seminal paper "Computing Machinery and Intelligence": "Can machines think?" This question has since become foundational in AI, inspiring decades of research and guiding ethical discussions around AI's role in society. Turing introduced the Turing Test, a benchmark to evaluate a machine's ability to exhibit human-like intelligence by convincingly mimicking human responses. Though the concept was theoretical, it has influenced both the practical goals and ethical considerations of AI development, particularly in fields like critical infrastructure where human-like decision-making capabilities are increasingly sought[27].

The formalization of AI as a field occurred in 1956 during the Dartmouth Conference, where researchers such as John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon discussed the potential of machines capable of performing tasks associated with human intelligence[28]. This meeting, often regarded as the "birth" of AI, not only introduced the term "artificial intelligence" but also laid down foundational ideas that would steer AI research for decades. These principles have since evolved to address the specific requirements of critical infrastructure systems, which require AI solutions capable of autonomous decision-making and robust risk management.

In the 1950s and 1960s, early AI research focused on symbolic AI, where machines were programmed to manipulate symbols and apply logical rules to solve problems. Groundbreaking programs such as Allen Newell and Herbert A. Simon's Logic Theorist[29], which could prove mathematical theorems, and the General Problem Solver (GPS), designed to mimic human problem-solving, demonstrated that machines could handle tasks requiring reasoning and planning. These early innovations laid the groundwork for the sophisticated, predictive AI systems that now support critical infrastructure, helping to optimize complex operations, prevent failures, and ensure resilience.

This chapter traces the historical and conceptual progression of AI in the context of critical infrastructure, leading to its current applications and future potential. Following this foundational overview, we will examine the development of machine learning, a key component that has enabled AI systems to analyze vast amounts of data and improve over time. We then explore notable industrial applications, highlighting how AI supports risk management, real-time monitoring, and predictive maintenance in critical infrastructure sectors. The chapter concludes with a look into future trends,

[26] This chapter was generated with the assistance of OpenAI's ChatGPT, an AI language model designed to facilitate and support the drafting process. While every effort has been made to ensure accuracy and relevance, the content has been reviewed and validated by the author to meet the specific objectives of this report. ChatGPT was utilized as a tool to enhance productivity and provide inspiration, and all final decisions and interpretations are those of the author.

[27] (Turing, 1950)

[28] (McCarthy, McCarthy, Minski, Rochester, & Shannon, 1955)

[29] (Newell & Simon, 1956)

considering emerging AI capabilities poised to address new challenges in safeguarding essential systems against evolving risks.

3.2 Expert Systems and Knowledge-Based Systems (1970 - …)

The 1970s saw the development of expert systems, which were designed to mimic the decision-making abilities of human experts. One of the most famous early expert systems was MYCIN[30], developed at Stanford University, which could diagnose bacterial infections and recommend treatments. MYCIN represented a significant advancement in the application of AI to real-world problems, demonstrating the potential of AI in critical areas like healthcare. These systems used knowledge bases of facts and rules to make inferences and provide recommendations, highlighting the practical utility of AI. The 1980s witnessed a boom in the development and commercialization of expert systems. These systems were deployed across various industries, including energy, where they helped optimize power generation and distribution. By leveraging AI, these expert systems provided significant improvements in operational efficiency and decision-making. For instance, in the energy sector, AI-driven systems optimized the scheduling of power generation and maintenance activities, leading to cost savings and enhanced reliability.

Energy Management

Early applications of AI in energy systems involved automating control systems in power plants to enhance operational efficiency and reliability. AI techniques monitored and controlled various aspects of power generation and distribution, contributing to more stable and efficient energy management[31]

ICT and Power Systems

The interplay between information and communication technologies (ICT) and power systems introduced new security challenges. Hadjsaid et al. (1997) emphasized the need for research to address ICT vulnerabilities in power systems, highlighting early uses of AI to enhance cybersecurity measures and protect critical infrastructure[32].

Predictive Maintenance

In the late 1980s, AI began to be employed for predictive maintenance, using data and AI algorithms to predict equipment failures. This approach reduced downtime, maintenance costs, and improved overall reliability of critical infrastructure. Early models developed by IBM's Watson Research Center were among the pioneers in this field[33].

Quality Control and Defect Detection

AI technologies enhanced quality control processes in manufacturing during the late 1980s. Machine learning algorithms analyzed product images to detect defects with higher accuracy than human inspectors, significantly improving detection rates in industries like electronics manufacturing[34].

Robotics and Automation

[30] (Buchanan & Shortliffe, 1984)
[31] (Power, 1991)
[32] (Hadjsaid, Sabonnadière, & Canard, 1997)
[33] (Watson, 1988)
[34] (Smith & Jones, 1989)

The 1980s also saw the adoption of AI-driven robotics in manufacturing. These robots performed tasks ranging from simple assembly to complex welding and painting, increasing production speed and precision. The automotive industry was a key adopter, revolutionizing assembly line processes with AI-driven robotics[35].

Process Optimization

AI was used to optimize industrial processes in the late 1980s. Advanced algorithms analyzed vast amounts of data to identify inefficiencies and suggest improvements, leading to significant cost savings and improved production rates. General Electric, for example, reported notable increases in efficiency from AI-driven process optimization[36].

Intelligent Training Systems

Intelligent training systems (ITS) were among the first AI applications in industry, customizing training programs for employees based on their learning pace and skill levels. By the 1980s, ITS were being developed to improve worker training and efficiency, simulating complex industrial environments for safer and more effective training[37].

3.3 Machine Learning and Statistical Approaches (1990s-2010s)

Support Vector Machines and Reinforcement Learning

The 1990s and 2000s saw significant advancements in machine learning, with the development of new algorithms such as Support Vector Machines (SVMs) and reinforcement learning. SVMs provided powerful tools for classification and regression tasks, while reinforcement learning enabled machines to learn from trial and error, making it particularly useful for decision-making in dynamic environments.

Deep Learning

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), in the early 2010s marked a major milestone in AI. Deep learning techniques, which involve training large neural networks with many layers, revolutionized fields such as image and speech recognition, as noted by Yann LeCun and colleagues[38]. This period also saw the application of deep learning to various aspects of critical infrastructure, including predictive analytics and anomaly detection.

Vulnerabilities in Power Systems

The increased complexity and interdependence of critical infrastructure brought new vulnerabilities, particularly in power systems. Stefanini and Ciapessoni (2001) emphasized the criticality of AI in assessing and managing these vulnerabilities[39]. AI was employed to analyze and mitigate risks associated with interconnected systems, ensuring their resilience and stability.

[35] (Klein, 1987)
[36] (GE Research, 1989)
[37] (McDermott, 1982) and (Bertin, Buciol, & Stefanini, 1998)
[38] (LeCun, 2015)
[39] (Stefanini & Ciapessoni, La vulnerabilità del sistema elettrico come infrastruttura interdipendente, Nov. 2001)

3.4 Big Data and Computational Power (2010s)

Smart Grids

The integration of AI into smart grids marked a significant leap in energy systems. AI technologies enhanced the management of electricity flow, improved fault detection, and facilitated the incorporation of renewable energy sources. The U.S. Department of Energy highlighted the importance of AI in this context: "AI in smart grids enhances real-time monitoring and fault detection."[40]

Energy Storage Management

AI plays a crucial role in optimizing the use of energy storage systems. By analyzing data on energy consumption and production, AI systems could efficiently match supply with demand, thereby improving the overall efficiency and reliability of energy systems. A report by the European Commission noted that "AI-driven energy management systems are crucial for the integration of renewables."[41]

ICT and Power Systems

The interplay between information and communication technologies (ICT) and power systems introduced new security challenges. Hadjsaid et al. (2007) discussed the "ICT vulnerabilities of power systems,"[42] emphasizing the need for a roadmap for future research to address these emerging challenges. AI was employed to enhance cybersecurity measures, ensuring the protection of critical infrastructure against cyber threats.

3.5 Rise of Large Language Models (LLMs) (2018-Present)

Transformers

The introduction of transformers, as described in the 2017 paper "Attention is All You Need," [43]revolutionized natural language processing (NLP). Transformers enabled the development of large language models (LLMs) that could understand and generate human-like text with unprecedented accuracy and coherence.

BERT and GPT Series

Models like BERT (2018)[44] and OpenAI's GPT series (GPT-3 in 2020) demonstrated the power of large-scale pre-training and fine-tuning. These models set new benchmarks in NLP, enabling a wide range of applications from automated customer service to advanced research. According to OpenAI, their tool allows to give…marketing teams at companies across industries a better understanding of their customers' wants and needs[45].

[40] (DOE, US Dept. of Energy, 2019)
[41] (European Commission, 2024)
[42] (Hadjsaid, et al., 2007)
[43] (Vashvani, et al., 2017)
[44] (Devlin, Ming-Wei, Lee, & Toutanova, 2019)
[45] March 25, 2021: GPT-3 powers the next generation of apps. Over 300 applications are delivering GPT-3–powered search, conversation, text completion, and other advanced AI features through our API. https://openai.com/index/gpt-3-apps/

AI and Critical Infrastructure Protection

AI technologies have become integral to the cybersecurity of critical infrastructure. AI systems can detect and respond to threats in real-time, ensuring the continuous and secure operation of essential services. On the other hand, Generative AI may also induce risks, as stated by the US NIST[46].

3.6 Current Trends and Technologies

IoT and AI Integration

The integration of the Internet of Things (IoT) and Artificial Intelligence (AI) is transforming the management of critical infrastructure. IoT devices generate vast amounts of real-time data, which AI algorithms analyze to optimize performance, enhance operational efficiency, and detect anomalies. This synergy is particularly impactful in the energy sector, where AI processes data from IoT sensors to enable predictive maintenance, efficient energy usage, and improved decision-making. According to a Siemens publication *'AI can help us to make data-driven decisions quickly for man contingencies. For example, to react to critical situations, like a strong in-feed from renewables or a system fault, AI trained by simulations and analysis of hundreds of eventualities can rapidly generate the right countermeasures to keep systems running safely. In smart grids, AI manages electricity flow in real-time, balancing supply and demand and integrating renewable energy sources more effectively'*[47]. The U.S. Department of Energy highlighted that *"AI in smart grids enhances real-time monitoring and fault detection,"* contributing to grid stability.[48]

AI in Nuclear Power

AI applications in the nuclear power sector include reactor control, fault diagnosis, and radiation monitoring. AI-driven systems analyze vast amounts of reactor data to maintain optimal operating conditions and ensure safety. AI algorithms rapidly diagnose faults, reducing downtime and enhancing continuous operation. The International Atomic Energy Agency (IAEA) stated, "*Artificial intelligence methods…can similarly accelerate the fields of nuclear applications, science, and technology toward the IAEA goals of contributing to peace, health, and prosperity.*[49]" AI also improves radiation monitoring, providing early warnings of potential leaks or breaches, thus protecting plant workers and surrounding communities.

Smart Grids and Energy Storage Management

AI plays a crucial role in the development of smart grids by managing electricity flow, integrating renewable energy sources, and enhancing grid reliability. AI systems optimize energy storage solutions, such as batteries, by predicting energy demand and supply patterns, ensuring efficient use of stored energy. This optimization helps stabilize the grid and balance energy fluctuations. The European Commission noted, "*AI-driven energy management systems are crucial for the integration of renewables.*[50]"

AI in Transportation Infrastructure

[46] On April 29, 2024, NIST released a draft publication based on the AI Risk Management Framework (AI RMF) to help manage the risk of. The draft AI RMF Generative AI Profile can help organizations identify unique risks posed by generative AI and proposes actions for generative AI risk management that best aligns with their goals and priorities.
[47] Next-Gen Industrial AI - Energy Sector, https://assets.new.siemens.com/siemens/assets/api/uuid:fef90d09-6876-4510-b29b-bb6d60374793/siemens-next-gen-industrial-ai-energy-sector.pdf
[48] (US Dept. of Energy, 2024 (April))
[49] (IAEA - Int. Atomic Energy Agency, 2022)
[50] (European Commission, 20 June 2024)

AI enhances transportation infrastructure by optimizing traffic flow, reducing congestion, and improving safety. AI-driven traffic management systems predict traffic patterns, adjust signal timings, and reroute traffic. In railways, AI enables predictive maintenance of tracks and trains, improving safety and reliability. The National Transportation Safety Board is calling on the Federal Railroad Administration to formulate a plan to incorporate promising new technology into the existing system that prevents certain train collisions.[51]

AI in Water Supply Management

AI technologies improve water supply systems by monitoring water quality, managing distribution networks, and predicting demand patterns. AI-driven systems detect anomalies and predict maintenance needs, reducing water loss and ensuring efficient distribution. The World Bank reported, "AI-driven water management systems are essential for ensuring the sustainability and reliability of urban water supplies[52].

AI in Cybersecurity

As critical infrastructure become more interconnected, AI-based cybersecurity solutions are essential for detecting and responding to threats in real-time. AI algorithms analyze network traffic, identify unusual patterns, and detect potential security breaches before significant damage occurs. These systems also automate responses to cyber threats, enhancing overall security. According to the National Institute of Standards and Technology (NIST) AI-based cybersecurity solutions are pivotal for protecting critical infrastructure. Subsequently, on April 29, 2024, *"NIST released a draft publication based on the AI Risk Management Framework to help manage the risk of Generative AI.*[53]"

In summary, the integration of IoT and AI, advancements in nuclear power safety, the development of smart grids, energy storage management, AI in transportation and water supply management, and AI-driven cybersecurity are key trends shaping the future of critical infrastructure. These innovations enhance efficiency, safety, and reliability, showcasing the transformative potential of AI across various sectors.

3.7 Future Directions and Challenges

Sustainability and AI

Artificial Intelligence (AI) is increasingly vital in advancing sustainability, particularly in energy systems. AI technologies optimize energy consumption, enhance efficiency, and facilitate the integration of renewable energy sources into the grid. For instance, AI algorithms manage smart grids by balancing supply and demand, predicting energy needs, and improving the operation of renewable sources like wind and solar. The International Energy Agency (IEA) underscored the importance of AI in this context, stating*, 'Artificial intelligence (AI) holds promising potential for advancing nuclear energy production'*[54]. By utilizing predictive analytics and real-time data, AI helps reduce energy waste and promotes the adoption of cleaner energy technologies, significantly contributing to global sustainability targets.

AI-driven systems also enhance energy storage solutions, allowing for better management of renewable energy resources and reducing reliance on fossil fuels. For example, AI technologies

[51] (NTSB - Nat. Transportation Safety Board, 2023)
[52] The future of water: How innovations will advance water sustainability and resilience worldwide (World Bank, 2020)
[53] (NIST - Nat. Institute for Standard and Technology, 2024)
[54] (Picot, 2023)

optimize battery storage systems by predicting energy storage needs and adjusting operational parameters accordingly, which helps in stabilizing the grid and integrating intermittent renewable sources. The deployment of AI in smart energy systems exemplifies how technological advancements can address environmental challenges and support long-term sustainability goals.

Ethical and Regulatory Considerations

As AI becomes more integrated into critical infrastructure, addressing ethical and regulatory concerns is paramount. The deployment of AI raises significant issues related to data privacy, security, and fairness. Ensuring the responsible use of AI involves developing comprehensive ethical guidelines and regulatory frameworks that protect individual rights while promoting technological advancement. The World Economic Forum has highlighted the importance of these considerations, noting, "*Artificial intelligence (AI) holds promising potential for advancing nuclear energy production.*"[55] AI systems often rely on large datasets, raising concerns about data privacy and the potential for misuse. Establishing clear regulations to govern data collection, storage, and usage is essential to safeguard against privacy breaches and ensure transparency. Additionally, addressing algorithmic bias and ensuring fairness in AI decision-making processes are critical to preventing discrimination and ensuring equitable outcomes. Developing robust regulatory frameworks will help manage these risks and promote trust in AI technologies.

Generative AI

Generative AI, which involves creating new data or solutions based on existing information, is transforming various domains, including critical infrastructure management. This technology enables the development of innovative solutions and enhances the capabilities of AI systems in areas such as design, simulation, and predictive analytics. Stefanini (2024) emphasizes the impact of generative AI, stating, "*Generative transformational artificial intelligence is revolutionizing how we approach problem-solving in various domains*"[56].

In critical infrastructure, generative AI can be used to model complex systems, simulate different scenarios, and develop new strategies for optimization and risk management. For example, in energy systems, generative AI models can predict potential system failures, design more efficient grid layouts, and create advanced control algorithms. This technology not only improves operational efficiency but also enhances the ability to respond to emerging challenges and disruptions. The potential of generative AI extends beyond current applications, offering opportunities for significant advancements in infrastructure resilience, innovation, and performance. As the technology evolves, it will continue to play a crucial role in shaping the future of critical infrastructure.

In conclusion, while AI offers transformative benefits for sustainability, ethical governance, and innovation, addressing these future directions and challenges is essential for maximizing its potential and ensuring its responsible deployment. As we advance, a balanced approach to integrating AI into critical infrastructure will be crucial for achieving long-term success and resilience.

3.8 Notable Projects and Initiatives

Grid Modernization Initiative (GMI)

The U.S. Department of Energy's Grid Modernization Initiative (GMI) focuses on integrating AI technologies to enhance the reliability, resilience, and efficiency of the electrical grid. AI-driven

[55] (World Economic Forum, 2020)
[56] (Stefanini, Analysis and Application of Generative Transformational Artificial Intelligence: two case studies on Chat GPT, 2024)

enhancements in real-time monitoring, fault detection, and energy management are crucial. AI systems predict and mitigate potential failures, manage electricity distribution, and integrate renewable energy sources. The U.S. Department of Energy has noted that "*AI in smart grids enhances real-time monitoring and fault detection,*"[57] which is essential for maintaining grid stability and efficiency. Under GMI, the Smart Grid Investment Grant Program aims to deploy smart meters and advanced grid technologies. These efforts improve the grid's capacity to handle disruptions and ensure a continuous power supply.

European Smart Grids and Smart Cities

The European Union funds projects under programs like Horizon 2020 to leverage AI for developing smart grids and smart cities. These initiatives integrate renewable energy sources, optimize energy consumption, and enhance urban sustainability. The EU's INTERFLEX project, for instance, focuses on integrating distributed energy resources and enhancing grid flexibility through AI technologies[58]. Elena Ragazzi and Alberto Stefanini (2019) discussed the importance of robust security measures in these initiatives, stating, *"Are security standards for electricity infrastructure a good choice for Europe? Evidence on cost and benefits from two case studies."*[59] These projects demonstrate how AI can improve operational efficiency, reduce carbon footprints, and ensure sustainable urban growth. According to the European Commission, *"AI-driven energy management systems are crucial for the integration of renewables."*[60]

Google's DeepMind and the UK's National Grid

Google's DeepMind has partnered with the UK's National Grid to optimize energy use in data centers, significantly reducing carbon emissions and improving energy efficiency[61]. Advanced AI algorithms dynamically manage energy consumption, ensuring optimal data center operations. This collaboration underscores AI's potential to drive significant improvements in energy management and sustainability. The initiative has resulted in up to a 40% reduction in energy consumption at Google's data centers, illustrating the substantial impact of AI on efficiency.

Smart Energy Systems in Asia

Countries like China and Japan invest heavily in AI-driven smart energy systems. China's State Grid Corporation uses AI to enhance its power network's efficiency and reliability, predicting power demand and optimizing energy distribution. Japan's AI initiatives focus on disaster resilience, with systems designed to quickly restore power after natural disasters. A report by the International Energy "*assesses recent developments for over 50 components of the energy system that are critical for clean energy transitions. The components assessed include sectors, subsectors, technologies, infrastructure and cross-cutting strategies.*[62]"

AI-Driven Water Management Projects

AI technologies improve the efficiency and reliability of water supply systems. Singapore's Public Utilities Board (PUB) uses AI for real-time monitoring and predictive maintenance of its water distribution network. AI algorithms analyze sensor data to detect leaks and predict maintenance

[57] (US Dept. of Energy, 2024 (April))
[58] Interflex - Home (interflex-h2020.com)
[59] (Ragazzi & Stefanini, 2019)
[60] (European Commission, 2024)
[61] (ESO - Electricity System Operator UK National Grid, 2019)
[62] (IEA, 2023)

needs, reducing water loss and ensuring continuous supply[63]. According to the Smart Water Magazine, "*Digital transformation of the water sector* (is) *a game changer*.[64]"

Transportation Infrastructure Enhancements

AI is transforming transportation infrastructure by optimizing traffic flow and enhancing public transit systems. Cities like Los Angeles and New York use AI-driven traffic management systems to predict patterns, adjust signal timings, and reroute traffic to reduce congestion. These systems improve traffic flow, reduce emissions, and enhance urban mobility. The web site of World Economic Forum notes, "*Unmanned taxis are now available to book in Shanghai's suburban Jiading District*.[65]"

3.9 Recent Trends in AI Applications to Critical Infrastructure in Italy

Artificial Intelligence (AI) is transforming critical infrastructure across multiple sectors in Italy, particularly in energy, transport, and ICT. Each of these sectors is leveraging AI technologies to improve efficiency, reliability, and security[66].

In the energy sector, AI applications are playing a critical role in grid management, predictive maintenance, and the optimization of energy distribution. AI-driven models enhance grid reliability by predicting faults and enabling load balancing in real-time, while machine learning algorithms monitor systems such as turbines and transformers, identifying maintenance needs and optimizing performance. For example, AI integration in Italy's national electricity grid, as noted by Terna Group, enhances grid management by incorporating renewable energy sources efficiently and ensuring the continuous stability of the power supply[67].

The transport sector is also seeing a significant transformation through AI-driven traffic management systems, predictive maintenance, and the development of autonomous vehicles. AI-powered traffic management systems analyze patterns to alleviate congestion, while predictive maintenance reduces vehicle breakdowns and ensures the reliability of transport infrastructure. In Italy, AI is also central to smart mobility initiatives and the deployment of autonomous vehicles, which promises to enhance safety and operational efficiency, as highlighted by the Ministry of Infrastructure and Transport[68].

In the ICT sector, AI enhances cybersecurity, network optimization, and data center management. AI-based cybersecurity systems detect and respond to threats in real-time, protecting critical digital infrastructure. AI is also used to optimize telecommunications networks, improve data center energy efficiency, and forecast infrastructure needs to avoid bottlenecks. Initiatives led by Telecom Italia[69] and AGID[70] exemplify Italy's push towards AI-driven innovation in telecommunications and cybersecurity, ensuring a resilient digital infrastructure.

Overall, the adoption of AI across these sectors demonstrates its pivotal role in advancing critical infrastructure in Italy, ensuring their adaptability, security, and efficiency in the face of evolving demands.

[63] (PUB-Singapore's National Water Agency, 2020)
[64] Digital transformation of the water sector as a game changer (smartwatermagazine.com)
[65] (World Economic Forum, 2020)
[66] (ENEA, 2022)
[67] (TERNA group, 2023)
[68] (AGID, 2022)
[69] (Centro Studi TIM, 2023)
[70] (AGID, 2022)

3.10 Conclusion

Artificial intelligence (AI) has rapidly evolved into a foundational tool for enhancing the efficiency, security, and resilience of critical infrastructure worldwide. As the complexity and interdependencies of essential systems like energy, water, transport, and information and communication technologies (ICT) continue to expand, AI provides capabilities that are indispensable for managing these vast networks effectively. This review highlights the transformative potential AI brings to each domain, emphasizing its role in ensuring adaptability and resilience amidst ever-changing operational demands.

In the energy sector, AI has proven pivotal in optimizing smart grids, integrating renewable sources, and managing storage systems to maintain stability in power supply. Initiatives such as the U.S. Grid Modernization Initiative and partnerships between technology firms and national grids underscore how AI-driven advancements in real-time monitoring, fault detection, and distribution management are essential in maintaining grid reliability. These initiatives illustrate AI's capacity to stabilize and secure energy systems, even in the face of fluctuating demands and potential disruptions. Italy's national electricity grid exemplifies how AI is instrumental in load balancing, predictive maintenance, and overall grid management—an increasingly vital asset as renewables are further integrated. AI's deep integration into the energy sector exemplifies its indispensable role in maintaining resilient infrastructure and supporting the global transition to sustainable energy sources.

AI has also made significant strides in water and transportation infrastructure, where it enhances both efficiency and responsiveness. Real-time data analysis has streamlined urban water management systems, reduced water loss, and enabled predictive maintenance, as illustrated by Singapore's Public Utilities Board (PUB). In transportation, AI-driven solutions have transformed urban mobility through optimized traffic flow and autonomous systems that respond dynamically to changing conditions, reducing congestion and emissions. Cities like Los Angeles and New York demonstrate how smart traffic management systems mitigate congestion by analyzing patterns, adjusting signal timings, and rerouting traffic in real-time. Italy's focus on AI-enhanced transportation and smart mobility promises to elevate both safety and operational efficiency, aligning with a broader global trend of AI-optimized transport infrastructure.

In the ICT sector, AI plays a crucial role in safeguarding digital infrastructure through enhanced cybersecurity measures and network optimization. Italy's investments in telecom and cybersecurity exemplify how AI enhances the efficiency and resilience of digital infrastructure. As data and communications networks become integral to the functionality of all critical infrastructure, the application of AI-based cybersecurity systems that can detect, analyze, and respond to threats in real-time is essential to protect these interconnected systems. Globally, AI-driven cybersecurity measures are emerging as vital tools for protecting the digital backbone upon which most other critical services rely, highlighting the importance of resilient digital infrastructure as the lifeline of modern society.

However, while AI's role in optimizing and securing critical infrastructure is clear, its widespread deployment raises ethical, regulatory, and sustainability concerns that must be addressed to ensure responsible adoption. AI's reliance on vast datasets presents risks regarding data privacy, security, and fairness. As these systems permeate critical sectors, it is crucial to develop robust ethical guidelines and transparent regulatory frameworks to protect privacy, ensure fairness, and foster public trust. Sustainability considerations are also paramount, as AI technologies that optimize resource use must themselves be deployed in an energy-efficient manner. Striking a balance between advancing AI capabilities and minimizing their environmental impact will be essential in achieving global sustainability goals, especially in sectors like energy, where AI solutions significantly enhance the efficiency of renewable sources and storage systems.

Generative AI, a more recent advancement in AI technology, opens new opportunities for innovation in critical infrastructure, particularly in risk management, design, and predictive analytics. This technology enables the simulation of complex scenarios, the anticipation of potential failures, and the creation of efficient strategies for infrastructure planning and management. In energy systems, generative AI can model grid layouts, predict system vulnerabilities, and develop advanced control algorithms, further enhancing the resilience and adaptability of these systems. As this technology continues to advance, it promises to provide critical infrastructure with increased flexibility and innovation, thereby reinforcing its capacity to respond to a range of challenges, both foreseeable and unforeseen.

In conclusion, AI's integration into critical infrastructure represents not merely a technological progression but a transformative shift in the design, management, and protection of essential systems. The examples in this survey underscore the necessity of a strategic and forward-looking approach to AI deployment—one that upholds resilience, sustainability, and ethical principles. Addressing these considerations thoughtfully will allow AI to achieve its full potential as a tool for advancing societal good. Looking forward, ongoing investments in AI research, partnerships across sectors, and the development of policy frameworks will be fundamental in realizing AI's transformative impact on critical infrastructure globally, ensuring their sustainability and resilience in the face of future demands.

References

AGID. (2022, May 13). *Smarter Italy: il programma che spinge la smart mobility in Italia.* Tratto da Smarter Italy: con l'innovazione si cresce: https://smarteritaly.agid.gov.it/index.php/2022/05/13/smarter-italy-il-programma-che-spinge-la-smart-mobility-in-italia/

Bertin, A., Buciol, F., & Stefanini, A. (1998, Feb.). Towards Industrial Application of Intelligent Training Systems. *Expert Systems, 15*(1), p. 1-21. doi: https://doi.org/10.1111/1468-0394.00060

Buchanan, B. C., & Shortliffe, E. H. (1984). *Rule-Based Expert Systems: the Mycin Experiment of the Stanford Heuristic Programming Project.* Reading, MA: Addison Wesley. Retrieved 08 02, 2024, from https://people.dbmi.columbia.edu/~ehs7001/Buchanan-Shortliffe-1984/MYCIN%20Book.htm

Centro Studi TIM. (2023, Dec. 5). *L'intelligenza artificiale in italia: Mercato, Innovazione, Sviluppi.* Retrieved from Centro Studi TIM: https://www.gruppotim.it/content/dam/gt/centro-studi-tim/ai/final/05122023_Report%20AI_completo_fin.pdf

DOE, US Dept. of Energy. (2019). *AI in smart grids enhances real-time monitoring and fault detection.U.S. Department of Energy.*

ENEA. (2022, Aug. 09). *Energia: intelligenza artificiale a servizio di interventi di efficientamento sempre più efficaci.* Retrieved from MEDIA - Sito tematico ENEA: https://www.media.enea.it/comunicati-e-news/archivio-anni/anno-2022/energia-intelligenza-artificiale-a-servizio-di-interventi-di-efficientamento-sempre-piu-efficaci.html

ESO - Electricity System Operator UK National Grid. (2019, Aug. 21). *Former DeepMind expert's AI tool could help boost National Grid ESO's solar forecasts.* Retrieved from ESO: https://www.nationalgrideso.com/news/former-deepmind-experts-ai-tool-could-help-boost-national-grid-esos-solar-forecasts

European Commission. (20 June 2024, June 20). *Artificial Intelligence for Next Generation Energy.* doi:10.3030/101016508

European Commission. (2024, May 17). *Artificial Intelligence for Next Generation Energy*. Retrieved from CORDIS: https://cordis.europa.eu/article/id/451051-next-generation-energy-powered-by-artificial-intelligence

GE Research. (1989). Process Optimization through AI. *GE Research Journal, 45(3),* 233-240.

General Electric. (2022, April 4). *GE Using AI/ML to Reduce Wind Turbine Logistics and Installation Costs.* Tratto da GE Research: https://www.ge.com/research/newsroom/ge-using-aiml-reduce-wind-turbine-logistics-and-installation-costs#

Hadjsaid, N., Rognon, J., Caire, R., Stefanini, A., Flataboe, N., Ruzante, G., . . . Deconinck, G. (2007). *ICT Vulnerabilities of Power Systems: A Roadmap for Future Research.* Luxembourg: Office for the Official Publications of the European Commission.

Hadjsaid, N., Sabonnadière, J. C., & Canard, J. F. (1997). ICT Vulnerabilities in Power Systems: A Roadmap for Future Research. *Electric Power Systems Research, 41(2),* 91-100.

IAEA - Int. Atomic Energy Agency. (2022). *Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology.* Retrieved 08 05, 2024, from IAEA: https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology

IEA. (2023, July). *Tracking Clean Energy Progress 2023: Assessing critical energy technologies for global clean energy transitions.* Retrieved 08 04, 2024, from IEA: https://www.iea.org/reports/tracking-clean-energy-progress-2023

Keppelmann, J. (2024, Feb. 16). *Reducing defects and downtime with AI-enabled automated inspections.* Retrieved from IBM Newsletters: https://www.ibm.com/blog/reducing-defects-and-downtime-with-ai-enabled-automated-inspections/

Klein, L. (1987). AI and Robotics in Automotive Manufacturing. ,,. *Journal of Manufacturing Systems 6(2),* 123-132.

LeCun, Y. B. (2015). Deep learning. . *Nature, 521(7553),* 436-444.

McCarthy, McCarthy, J., Minski, M., Rochester, N., & Shannon, C. E. (1955). *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* Retrieved 08 02, 2024, from https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf

McDermott, J. (1982). R1: The Formative Years. . *AI Magazine, 3(4),* 21-32.

McKinsey & Co. (2023, April 12). *Maintenance and operations: Is asset productivity broken?* Retrieved from McKinsey & Company: Electric Power & Natural Gas: https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/maintenance-and-operations-is-asset-productivity-broken

Newell, A., & Simon, H. A. (1956). *The Logic Theory Machine: a Complex Information Process.* Retrieved 08 02, 2024, from http://shelf1.library.cmu.edu/IMLS/MindModels/logictheorymachine.pdf

Newell, A., Shaw, J. C., & Simon, H. A. (1959). *Report on a General Problem-solving Program.* Retrieved 08 02, 2024, from http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf

NIST - Nat. Institute for Standard and Technology. (2024, Apr. 30). *AI Risk Management Framework.* Retrieved from NIST: https://www.nist.gov/itl/ai-risk-management-framework

NTSB - Nat. Transportation Safety Board. (2023, Nov. 1). *Beyond Positive Train Control: Using New and Emerging Technologies to Improve Rail Safety.* Retrieved from NTSB: https://www.ntsb.gov/news/press-releases/Pages/NR20231101.aspx

Picot, W. (2023). Enhancing Nuclear Power Production with Artificial Intelligence. *IAEA Bullettin (Vol. 64-3)*. Retrieved 08 04, 2024, from https://www.iaea.org/bulletin/enhancing-nuclear-power-production-with-artificial-intelligence

Power, R. J. (1991). AI Applications in Power Grid Management. *Energy Journal, 12(4)*, 345-359.

PUB-Singapore's National Water Agency. (2020). *Smart water management – the Singapore's experience.* Retrieved from PUB-Singapore's National Water Agency: https://events.development.asia/system/files/materials/2020/11/202011-smart-water-management-singapore-experience.pdf

Ragazzi, E., & Stefanini, A. (2019). Are security standards for electricity infrastructure a good choice for Europe? Evidence on cost and benefits from two case studies. *International Journal of Critical InfrastructureVol. 15, No. 3, 15*(2). Retrieved 08 04, 2024, from https://www.inderscienceonline.com/doi/abs/10.1504/IJCIS.2019.100425?journalCode=ijcis

Smith, A. R., & Jones, T. S. (1989). AI in Quality Control: Case Studies from the Electronics Industry. *IEEE Transactions on Industrial Electronics, 36(3)*, 317-321.

Stefanini, A. (2024). *Analysis and Application of Generative Transformational Artificial Intelligence: two case studies on Chat GPT.* Torino: IRCRES. Retrieved 08 04, 2024, from https://www.ircres.cnr.it/wp-content/uploads/2024/01/RT_15_2024.pdf

Stefanini, A., & Ciapessoni, E. (Nov. 2001). La vulnerabilità del sistema elettrico come infrastruttura interdipendente. *AEI - Rivista Ufficiale dell'Associazione Elettrotecnica Italiana, 88*.

TERNA group. (2023, March 15). *TERNA: 2023 development plan for the national electricity.* Retrieved from TERNA Driving Energy Media: https://www.terna.it/en/media/press-releases/detail/2023-development-plan

Turing, A. (1950). Computing Machinery and Intelligence. *Mind, 59(236)*, 433-460. Retrieved 08 02, 2024, from https://academic.oup.com/mind/article/LIX/236/433/986238

US Dept. of Energy. (2024 (April)). *AI for Energy: Opportunities for a Modern Grid and Clean Energy.* US DoE. Retrieved 08 04, 2024, from https://www.energy.gov/sites/default/files/2024-04/AI%20EO%20Report%20Section%205.2g%28i%29_043024.pdf

US Dept. of Energy. (2024 (April)). *AI for Energy: Opportunities for a Modern World Economy.* DoE. Retrieved 08 04, 2024, from https://www.energy.gov/sites/default/files/2024-04/AI%20EO%20Report%20Section%205.2g%28i%29_043024.pdf

Vashvani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017, Jun 12). *Attention is All You Need.* Retrieved from ArXiV: https://arxiv.org/abs/1706.03762

Watson, G. (1988). Predictive Maintenance Models at IBM. . *IBM Journal of Research and Development, 32(2)*, 213-222.

World Bank. (2020, June 15). *The future of water: How innovations will advance water sustainability and resilience worldwide.* Retrieved from worldbank.org: https://blogs.worldbank.org/en/water/future-water-how-innovations-will-advance-water-sustainability-and-resilience-worldwide

World Economic Forum. (2020, Jan. 14). *How global tech companies can champion ethical AI*. Retrieved from World Economic Forum: https://www.weforum.org/agenda/2020/01/tech-companies-ethics-responsible-ai-microsoft/

World Economic Forum. (2020, July 13). *These driverless cars in Shanghai form the world's first 'robotaxi' fleet.* Retrieved from World Economic Forum: https://www.weforum.org/agenda/2020/07/autonomous-vehicles-taxi-mobility/

4 AI GOVERNANCE FRAMEWORKS AND MODELS SUPPORTING CRITICAL INFRASTRUCTURE RESILIENCE *(Luigi Carrozzi, Alberto Stefanini)*

The introduction of artificial intelligence systems poses significant challenges from an organizational point of view. The implementation of these systems involves a considerable effort by the organization in order to ensure that the AI solutions introduced provide the maximum potential and do not negatively impact operational processes, in particular the "mission critical" ones, thus affecting, for example, Critical Infrastructure' provision of essential services. This chapter introduces the relevant case of bias management and presents solutions to properly manage those challenges adopting suited governance approaches.

4.1 Artificial Intelligence in mission critical context. Organizational challenges and the case of bias management. *(Luigi Carrozzi)*

The adoption of AI technology within an organization may involve a profound review of its operations processes, requiring new competences, organizational roles, and review of standard risk management, compliance and control practices. In this sense it is essential to identify what the new posture of the organization should be in order to achieve an effective and responsible control of the AI artifacts exploiting all the advantages they may generate, but, at the same time being fully accountable on what is the value at stake and adopting a wise, farsighted and responsible AI management approach. Let's take for example the problem of Bias. We may be reasonably aware that the level of autonomy and consequently the level trust we give to the AI systems is a key factor. But AI outputs may be affected by biases differently originated.

Bias management is one of the main problems an organization has to face when adopting AI solutions, and especially when mission critical functions are at the stake. The human (typically domain's experts) and the organizational knowledge management process is largely involved, that is, organization should be aware of possible consequences of biased output and manage this issue consequently.

These concerns are obviously greater when it comes to using such technology in mission critical contexts. As represented in the NIST Special Publication 1270[71] the harmful effects of AI bias are mainly focused on the representativeness of data and the fairness of machine learning algorithms, but as well human and systemic institutional and societal factors are significant sources of AI bias and should be properly considered.

And is a fact that when adopting AI systems, the organization as whole (as typically made up of, people, organizational roles, processes and technology) should be involved adopting a sound change management process to seize the opportunities and adequately manage the related risks. As highlighted by NIST in his publication:

[71] National Institute of Standards and Technology. (2022). *NIST Special Publication 1270 - Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*

"…It is also important to note that governance does not simply focus on technical artifacts, such as AI systems alone, but also on organizational processes and cultural competencies that directly impact the individuals involved in training, deploying and monitoring such systems."

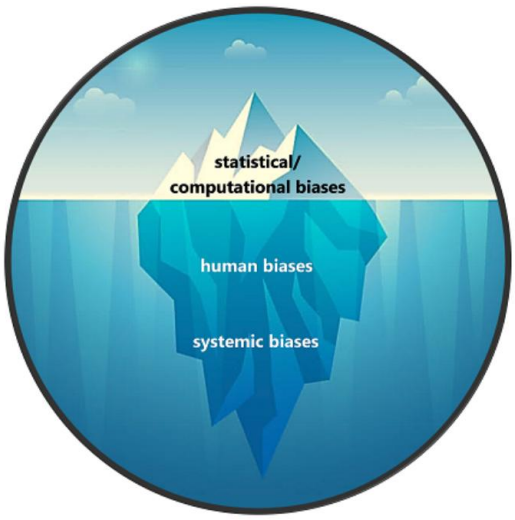


Fig. 4.1 The challenge of managing AI bias

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

The aim of NIST document has the objective to provide "socio-technical guidance" for identifying and managing AI bias. Specifically:

1. describes the stakes and challenge of bias in artificial intelligence and provides examples of how and why it can chip away at public trust;
2. identifies three categories of bias in AI - systemic, statistical, and human - and describes how and where they contribute to harms;
3. describes three broad challenges for mitigating bias - datasets, testing and evaluation, and human factors - and introduces preliminary guidance for addressing them.

So, the complex and multifaceted nature of bias of AI Systems may require the set-up of specific knowledge creation processes within the organization for the monitoring and analysis, on a continuous base, of possible biased output of the system, being able to prevent possible malfunctions/incidents and, in case this may occur, be prepared to put in place the related remedies. And this entails specific processes, procedures and domain-specific competences.

The case of acquisition of bias management capabilities is only one aspect of the new organizational posture that organizations, and especially those one performing critical functions, need to adopt to cope with when running AI assisted business operations.

The highly recommended solution of *Humans-in-The-Loop*" for a human-centered, responsible and trustworthy development and use of AI inevitably may involve a significative organization effort and new *cross-functions* and *cross-cultural* competences to best enable safe, ethic, trustworthy and business effective outputs by AI systems.

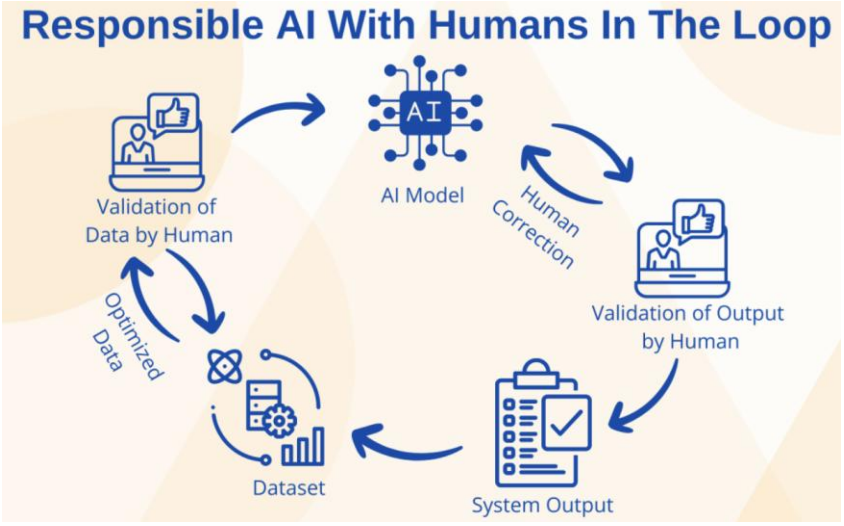


Fig. 4.2 Responsible AI use with Humans-in-the-Loop (HITL)

(Source: Anderson Anthony. (2023). Responsible AI use with Human-in-the-Loop (HITL) https://www.linkedin.com/pulse/responsible-ai-use-human-in-the-loop-hitl-anderson-anthony)

It is a fact that AI systems are designed, developed and used by humans. And humans apply to the construction of AI artifact with their capacities but also with their limits.

Controlling the sources of Human Bias, for example, has an overwhelming impact on the trustworthiness of AI systems.

In the following picture the potential impact of human bias is represented along the entire lifecycle of artificial intelligence.

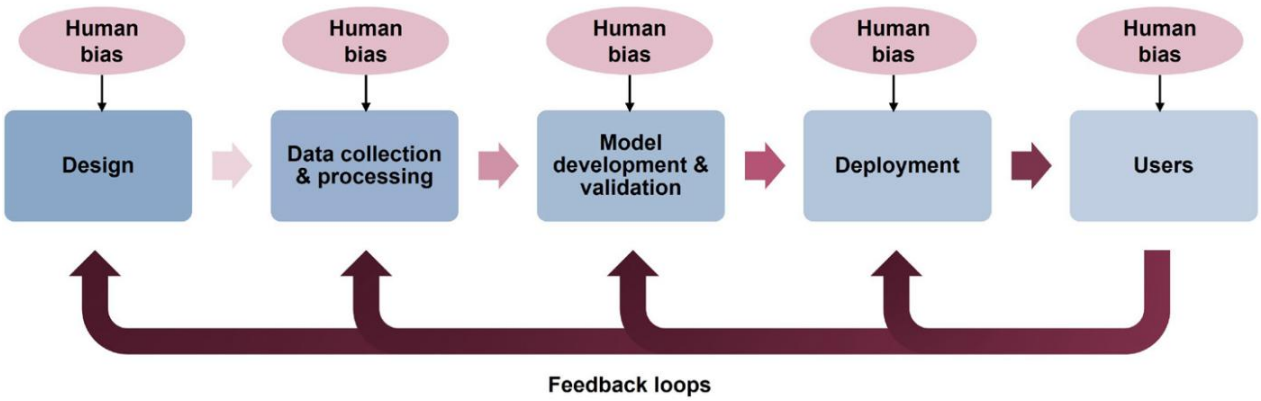


Fig. 4.3 Human bias in the artificial intelligence life cycle

(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)

Moreover, in each stage of the lifecycle of AI system, from design to the final stage of deployment and use, there are different types and sources of Bias as represented in picture 4.4.

It's clear that each actor involved in the production chain and use of AI system should be aware of the different types and sources of bias.

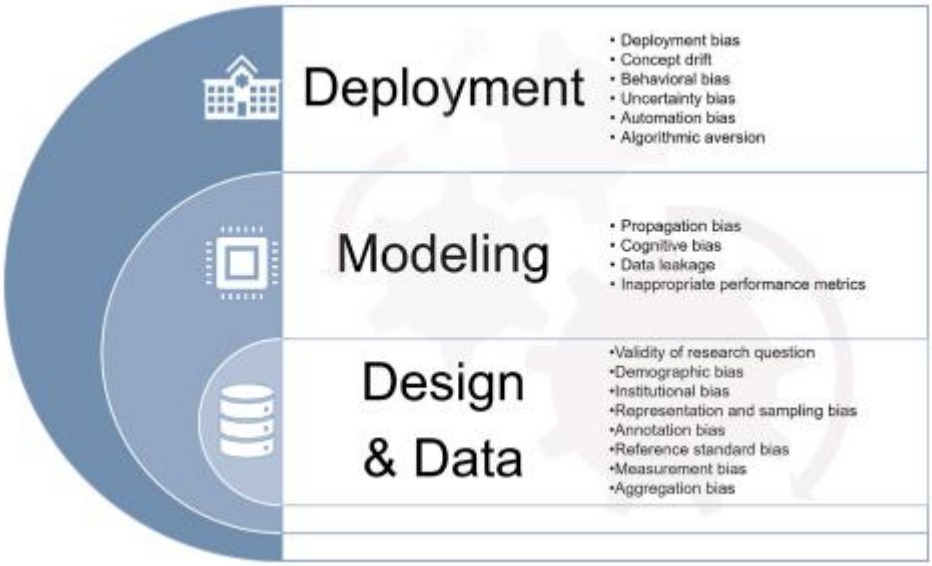


Fig.4.4. Main types and sources of bias

(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)

4.2 Organizational Impact of AI Adoption in Critical Infrastructure

(Alberto Stefanini)[72]

The adoption of Artificial Intelligence (AI) systems within critical infrastructure (CI) represents not only a technological transformation but also a significant organizational challenge. As AI becomes integrated into core functions of energy grids, telecommunications, transport, and healthcare systems, the organization itself must evolve to ensure security and resilience. Below, we will analyze two key areas: (a) the role of the organization in ensuring security and resilience of CI with AI systems; and (b) the role of top management in setting strategic directions to support AI-driven security and resilience.

(a) The Role of the Organization in Ensuring Security and Resilience of CI with AI Systems

AI adoption in CI redefines the roles, responsibilities, and structures of organizations. Instead of treating AI systems as discrete technical assets, organizations need to treat them as integral parts of broader operational, risk management, and governance frameworks. The organization's role in securing AI-enabled CI services is multi-dimensional.

- **Cross-Functional Collaboration.** Effective AI governance requires close collaboration between diverse functions within the organization, such as IT, legal, compliance, risk management, operations, and human resources. The complexity of AI adoption demands an interdisciplinary approach where each department understands its role in maintaining security and resilience. For example, legal teams must ensure compliance with data protection laws, while IT and risk management must work together to address the cybersecurity risks of AI systems.
- **Organizational Change Management.** AI deployment in CI often entails re-engineering existing processes, workflows, and reporting structures. Organizations must anticipate and manage these changes by implementing structured change management strategies. This involves educating employees, redefining job roles, and fostering a culture that embraces AI's benefits while recognizing its risks. Employees, especially those involved in mission-critical operations, must be adequately trained to understand AI's role and limitations, including what safeguards are in place to mitigate risks.
- **Resilience through Continuous Monitoring.** The ability to continuously monitor AI systems, evaluate their performance, and proactively detect issues is crucial for organizational resilience. AI systems are not static; they evolve over time through learning and adaptation. Therefore, organizations must establish monitoring frameworks that track AI behaviors, measure key performance indicators (KPIs), and alert human operators to potential deviations or vulnerabilities. This capacity for ongoing evaluation allows organizations to rapidly respond to incidents, maintain operational continuity, and prevent disruptions in CI services.
- **Incident Response and Crisis Management.** Even with robust safeguards, there is no guarantee that AI systems will function flawlessly. Therefore, organizations need to have an effective incident response strategy specifically designed for AI-driven failures. This includes defining escalation procedures, ensuring rapid communication between departments, and having dedicated response teams ready to intervene in case of AI malfunctions or attacks. By

[72] This section was generated with the assistance of OpenAI's ChatGPT, an AI language model designed to facilitate and support the drafting process. While every effort has been made to ensure accuracy and relevance, the content has been reviewed and validated by the author to meet the specific objectives of this report. ChatGPT was utilized as a tool to enhance productivity and provide inspiration, and all final decisions and interpretations are those of the author.

developing an AI-specific crisis management plan, organizations can minimize the impact of system failures on their CI services and ensure swift recovery.

(b) The Role of Top Management in Providing Strategic Directions

Top management plays a crucial role in shaping the organization's strategy for AI adoption, particularly when it concerns critical infrastructure. Their commitment is essential in ensuring that AI initiatives are not limited to technical execution but are aligned with the organization's overarching goals for security, resilience, and ethical responsibility. The key roles of top management in this context include:

- **Vision and Strategic Alignment.** AI governance begins at the top. Executives must clearly articulate a vision for how AI will be integrated into CI operations, setting the tone for its responsible use and ensuring it aligns with the organization's long-term objectives. Top management must ensure that AI systems enhance the organization's resilience and deliver reliable, secure services while meeting regulatory and ethical standards.
- **Resource Allocation and Investment.** AI adoption in CI requires significant investments, not just in technology but in human resources, training, and infrastructure. Top management must allocate sufficient resources to build and maintain secure and resilient AI systems. This includes funding for, cybersecurity enhancements, and continuous AI system monitoring and validation.
- **Risk Governance and Ethical Oversight.** AI introduces new ethical concerns, particularly in relation to decision-making in critical contexts where human lives and societal well-being are at stake. Top management must establish robust risk governance frameworks that incorporate ethical oversight. They should ensure the organization is prepared to handle the societal, legal, and reputational risks associated with AI in CI, committing to transparency and accountability.
- **Policy Making and Public Engagement.** Given the societal importance of CI, top management should actively engage with policymakers, regulators, and stakeholders to influence the broader regulatory environment around AI in critical infrastructure. By shaping policies that prioritize safety, security, and ethical considerations, executives can ensure that their AI governance framework aligns with evolving national and international standards.

Conclusion

In summary, the integration of AI into critical infrastructure requires a paradigm shift at the organizational level. Safeguards must be put in place to ensure secure and resilient operations, cross-functional collaboration must be fostered, and top management must take an active role in setting strategic directions for AI adoption. Organizations must not only implement technical solutions but also cultivate the human and cultural competencies necessary for effectively managing AI's impact on critical infrastructure services. Through proactive leadership and structured change management, organizations can harness the potential of AI while safeguarding the security and resilience of critical infrastructure.

4.3 Frameworks and models for the Governance of AI systems in Critical Infrastructure *(Luigi Carrozzi)*

The specific nature and functioning of AI systems entail a profound knowledge of the implications of this technology that may have deep impacts on business processes and potential third parties concerned (stakeholders/shareholders, providers, customers, citizens) involving the overall intent statement, culture and values of the organization.

This is why a wise and farsighted business approach to the adoption of AI artifacts may not involve only the management and the operational layers of the organization but needs a strong commitment of the top management requiring a relevant involvement the governing body.

And it goes without saying that infrastructure providing mission critical services, to guarantee their resilience, need top notch control and governance of AI systems that should be the stronger the greater the degree of autonomy is entrusted to the AI system adopted.

According to this principle it may be useful to adopt specific governance frameworks when adopting AI systems.

On this regard it's worth to mention ISO/IEC 38507[73] standard, that introduces the governance implications of the use of AI. In the foreword of the document, it's explicitly mentioned that "…*As with any powerful tool, the use of AI brings new risks and responsibilities that should be addressed by organizations that use it. AI is not inherently 'good' or 'evil', 'fair' or 'biased', 'ethical' or 'unethical' although its use can be or can seem to be so. The organization's purpose, ethics and other guidelines are reflected, either formally or informally, in its policies. This document examines both governance and organizational policies and their application and provides guidance to adapt these for the use of AI. The operational aspects of the policies are implemented through management….".* And addresses his applicability as follows:

"…*This document provides guidance for members of the* ***governing body*** *of an organization to enable and govern the use of Artificial Intelligence…, in order to ensure its* ***effective***, ***efficient*** *and* ***acceptable use within the organization.***

This document also provides guidance to a wider community, including:

— *executive managers;*

— *external businesses or technical specialists, such as legal or accounting specialists, retail or industrial associations, or professional bodies;*

— *public authorities and policymakers;*

— *internal and external service providers (including consultants);*

— *assessors and auditors.*

This document is applicable to the governance of current and future uses of AI as well as the implications of such use for the organization itself.

[73] ISO/IEC 38507. (2022). *Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations*

The organization should be well aware that AI is not a "traditional" technology. Having the objective to get an adequate control of IA and since the possible involvement of the whole organization may be necessary, it's important to have a strong commitment by the top management of the organization, "the governing body", to allocate resources and act accordingly with this fundamental objective. AI Governance key factors may involve, among others, the decision-making process enabled by the inference engine of AI systems that need to be fully under control of the organization, in particular when mission critical activities and the provision of essential services are at the stake. In this regard, the consequences of the quality of governance systems may have high consequences on the resilience of the organization, in particular when the operational activities are supported by AI systems.

Among the different AI governance frameworks available, it's also worth of note the "AIGA AI Governance Framework"[74]. The AIGA (Artificial Intelligence Governance and Auditing) framework includes a model "*The Hourglass Model*"[75] that puts the base to the overall structure of the AI Governance framework, an "*AI Governance Lifecycle*" where AI governance tasks are mapped to the OECD's AI system lifecycle framework[76] and a "*Practical to-do list*" consisting of 67 tasks to support organizational AI governance. In the following picture is represented the structure of the "Hourglass model".

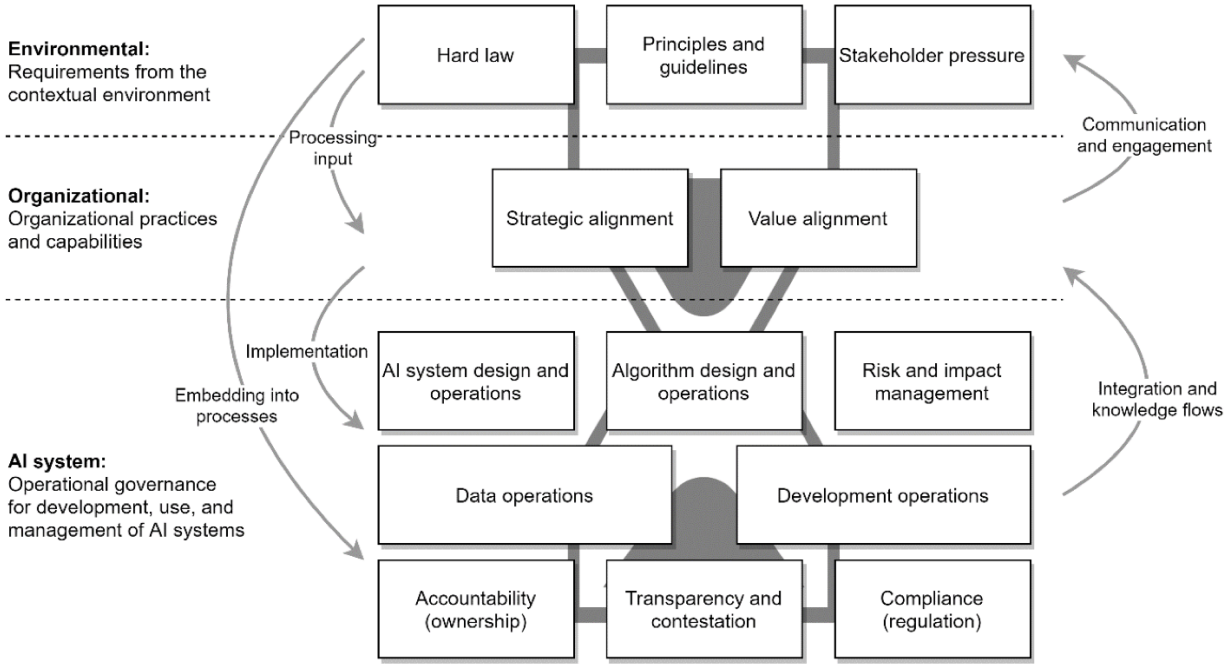


Fig. 4.5 The hourglass model of organizational AI governance

(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])

In this model, three layers are identified: the environmental layer, related to the requirements arising from the context; the organizational layer, that addresses the fundamental organizational practices and capabilities required; the AI System layer, where are stated the operational governance directions

[74] https://ai-governance.eu
[75] 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] https://doi.org/10.48550/arXiv.2206.00335
[76] OECD (2022). *OECD Framework for the Classification of AI systems*. OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en.

for the development, use and management of AI System. According to the description of the authors of the model "*...The hourglass metaphor denotes the flow of governance requirements from the environmental layer to AI systems through the mediating organizational layer. The metaphor also highlights the dynamic nature of AI governance as a continuous activity that translates the normative regulatory, self-regulatory, and stakeholder inputs into operational practices*".

By the fact, the AIGA AI Governance Framework should be considered as a robust framework oriented to practical implementation of governance principles for an organization intending gaining an overall control of the AI system adoption process.

4.4 A multifaced Governance approach of AI in Critical Infrastructure *(Alberto Stefanini)*[77]

The governance of Artificial Intelligence (AI) in critical infrastructure demands a multi-faceted approach that not only ensures compliance with various industry standards but also leverages the capabilities of public-private partnerships to foster resilience and security. Critical infrastructure, such as energy grids, transportation networks, and healthcare systems, rely increasingly on AI to automate and optimize operations. However, the adoption of AI technologies in these contexts introduces new risks, such as algorithmic opacity, bias, and cybersecurity vulnerabilities, which must be effectively managed.

Public-Private Partnerships and AI Governance.

A growing body of literature and industry practice suggests that public-private partnerships (PPP) play a crucial role in mitigating these risks. By involving both governmental oversight and private sector innovation, PPPs help ensure that critical infrastructure are not entirely dependent on private actors for security and governance. In contexts like AI governance, PPPs can bridge gaps in regulation, innovation, and best practices, providing a mechanism for knowledge transfer, risk-sharing, and collaborative response to cybersecurity threats.

The 2005 article, "Humans as a Critical Infrastructure," emphasizes the centrality of the human element in these infrastructure. Humans are not just operators; they are key nodes within the infrastructure, aggregating data, making real-time decisions, and interacting with AI systems in mission-critical environments. The concept of humans as a critical infrastructure highlights their vulnerability to cyberattacks, which could not only disrupt the individual but also the wider system they support. A successful cyberattack on such a human node could compromise the safety of people and the functionality of AI-powered systems they manage.

Modern public-private partnerships should be developed with this human factor in mind, ensuring that both technology and human operators are protected against potential threats. Governments and private entities need to collaborate to build frameworks that secure the human element, provide training, ensure resilience, and foster a comprehensive approach to AI governance. An example of

[77] This section was generated with the assistance of OpenAI's ChatGPT, an AI language model designed to facilitate and support the drafting process. While every effort has been made to ensure accuracy and relevance, the content has been reviewed and validated by the author to meet the specific objectives of this report. ChatGPT was utilized as a tool to enhance productivity and provide inspiration, and all final decisions and interpretations are those of the author.

such an approach is a shared V&V platform, which could be operated as a public-private initiative, ensuring that both public and private infrastructure adhere to safety and security standards while enabling rapid innovation.

Overview of standards supporting AI Governance

With respect to the due difference between Governance and Management, we should consider that the role competence of the board, related to overall direction setting, planning and oversight may be more effective where management functions recognize the multifaceted operational challenges of Artificial intelligence and adopt, preferably with an integrated and synergic approach, the proper combination of ISO, NIST, and sector-specific guidelines are necessary to support a robust governance. In this sense here we recall the previously mentioned ISO/IEC 38507: Governance of Artificial Intelligence whose purpose is to provide guidelines for organizations to govern AI systems responsibly, ensuring alignment with ethical principles and business objectives, to the aim of extending traditional governance frameworks, so as to accommodate AI's unique attributes, such as autonomy, complexity, and learning capabilities. In particular the key contribution of this standard is related to the promotion of top-management engagement, ensuring that AI governance is embraced at all decision-making levels and include ethical considerations, transparency, and accountability. Furthermore, in public-private partnerships, this standard can be a guiding document to ensure mutual accountability and the responsible deployment of AI in mission-critical systems.

However, as previously stated, governance functions may be largely supported by other specific standards, as those already mentioned in previous chapters, facing from the AI systems itself to other relevant practices such as Information Security, Risk Management and Critical Infrastructure Protection.

a) ISO/IEC 27001: Information Security Management System

The ISO/IEC 27000 Series: Information Security Management Systems (ISMS) provides a comprehensive framework for managing information security risks, including AI-related threats such as data breaches or cybersecurity vulnerabilities. The series is traditionally focused on IT security - in particular through ISO 27001[78]- and the associated standards, will be integrated (e.g. ISO/IEC 27090) for AI Cybersecurity

b) *The ISO/IEC 42001: Artificial Intelligence Management Systems (AIMS),*

The ISO/IEC 42001 standard is a newer framework designed to address the specific needs of managing AI systems within organizations. It emphasizes critical aspects such as risk management, compliance, and cybersecurity, particularly during the development and deployment of AI solutions. Unlike conventional IT frameworks, ISO/IEC 42001 tackles challenges unique to AI, including algorithmic bias, ethical risks, and system autonomy. This standard is especially valuable in critical infrastructure sectors, where it ensures that AI systems operate securely and adhere to predefined ethical and operational standards. It also plays a pivotal role in public-private partnerships by mandating that private sector AI systems integrated into public infrastructure meet stringent safety and ethical benchmarks.

c) The NIST Risk Management framework[79]

Similarly, the NIST Risk Management Framework (RMF) provides a structured, risk-based approach to managing cybersecurity risks in IT systems, including AI. Widely recognized in the United States, the RMF is crucial for securing mission-critical operations.

[78] (BSI 2004)
[79] (NIST – Nat. Institute for Standard and Technology, 2024)

d) the NERC CIP (Critical Infrastructure Protection) standards[80]

In the energy sector, the NERC CIP standards focus on safeguarding critical electrical infrastructure and are increasingly applicable to other domains like telecommunications and transport. These standards address risks associated with both human and automated systems, offering a comprehensive approach to securing critical infrastructure. By ensuring that AI systems do not introduce new vulnerabilities, NERC CIP plays a central role in maintaining the safety and reliability of mission-critical systems in various sectors.

Together, the above frameworks underscore the importance of robust governance, emphasizing security, compliance, and ethical considerations across the AI lifecycle in critical infrastructure.

Verification and Validation (V&V): the Human and AI Interface

In the governance of AI in critical infrastructure, Verification and Validation (V&V) play pivotal roles in ensuring the safety, reliability, and ethical compliance of AI systems. Public-private partnerships can benefit from shared V&V frameworks, enabling both sectors to test and certify AI systems against evolving threats and performance standards.

- **What should be verified?**

- The accuracy and fairness of AI outputs, especially in safety-critical tasks.

- The resilience of AI systems against cyberattacks, including both technical systems and the humans interacting with them.

- The compliance of AI with ethical standards, including bias prevention, transparency, and explainability.

- The alignment of AI systems with sector-specific regulations, ensuring that energy, transport, and healthcare systems remain secure and efficient.

- **What should be validated?**

- The operational integrity of AI systems in live environments, ensuring they meet predefined safety and security benchmarks.

- The performance of AI systems under various stress conditions, validating their ability to function effectively even during cyberattacks or operational disruptions.

- The effectiveness of human-AI collaboration, ensuring that human operators can make informed decisions and intervene when necessary.

- Why V&V is still important:

1. **Validation of Resilience:** The validation aspect ensures that AI systems meet their intended performance objectives, particularly in maintaining resilience during adverse scenarios, such as cybersecurity attacks or operational failures. It guarantees that the AI systems are functioning as expected in real-world conditions.

2. **Verification of Security Protocols:** Verification is crucial to ensure that all implemented security measures—whether they are for mitigating bias, data security, or other AI-specific risks—are in place

[80] (NERC, 2003-2009)

and functioning effectively. It also plays a role in meeting compliance with standards like ISO/IEC 27000 and NIST guidelines.

3. **AI-Specific V&V:** Unlike traditional systems, AI's learning capabilities introduce a dynamic component that needs continuous monitoring. V&V should account for these evolving characteristics, ensuring that initial safeguards continue to hold as the AI system adapts over time.

A Comprehensive Validation and Verification Platform for AI-Generated Software, Firmware, and Operational Procedures

This paragraph introduces the GALICIA platform, under the Galicia Project[81] exemplifying a comprehensive Validation and Verification (V&V) solution that aligns with the methodologies discussed in section 4.4. By addressing the challenges associated with AI-generated software, firmware, and operational procedures, it underscores the importance of robust V&V practices in ensuring reliability, security, and compliance in AI-driven innovations.

The GALICIA platform—Generative AI with Cybersecurity for Internet Applications Development—addresses the growing need for robust validation and verification (V&V) methodologies in AI-generated software, firmware, and operational procedures. It ensures reliability, security, and compliance across diverse applications by encompassing the entire lifecycle of AI-generated outputs, from development to deployment.

References

National Institute of Standards and Technology. (2022). NIST Special Publication 1270 -Towards a Standard for Identifying and Managing Bias in Artificial Intelligence

ISO/IEC 38507. (2022). Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations

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] https://doi.org/10.48550/arXiv.2206.00335

OECD (2022). OECD Framework for the Classification of AI systems. OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en.

ISO/IEC 27001 - Information Security Management System, https://www.bsigroup.com/en-GB/products-and-services/standards/iso-iec-27001-information-security-management-system/

BSI. (2004). *ISO/IEC 27001* - Information Security Management System, BSI: https://www.bsigroup.com/en-GB/products-and-services/standards/iso-iec-27001-information-security-management-system/

[81] GALICIA is a project funded by the European Union, within the framework of the NGI Sargasso. GALICIA aims to address the risks and challenges posed by Generative AI, particularly its impact on digital resilience. The project will focus on verifying the correctness and security of AI-generated code, ensuring compliance with user requirements and industry standards. By building trust in AI-generated software, GALICIA seeks to accelerate the adoption of AI in industrial automation while maintaining high levels of security and reliability.

ISO. (2022). ISO/IEC 15408 - Information technology -- Security techniques -- Evaluation criteria for IT security, https://www.iso.org/standard/72891.html

NERC. (2003-2009). *NERC CIP Solutions.* Atlanta, Georgia: North American Electric Reliability Corporation.

NIST - Nat. Institute for Standard and Technology. (2024, Apr. 30). *AI Risk Management Framework.* Retrieved from NIST: https://www.nist.gov/itl/ai-risk-management-framework

5 RISK ASSESSMENT AND MANAGEMENT *(Glauco Bertocchi, Francesca Della Mea, Giorgio Pizzi)*

Risk analysis is often a complex task that must consider various aspects (technical, legal, organizational, human, etc.) of the organization being examined. This task is certainly very complex in the case of Critical Infrastructure because of its importance for the well-being of society and the safety of people. The possibility of using Artificial Intelligence tools represents an opportunity for better managing the complexity of the risks but also the presence of new risks (many) arising from a new class of technologies. To better manage the risks of AI usage we will have to make use of the best of human intelligence to adopt the most effective methodologies.

5.1 Tools (standard, best practices, etc.) for defining and assessing AI risk

In this paragraph, we give some references to the most relevant standards, best practices, and international regulations for defining and assessing AI risk and its use for enhancement and management of resilience of the Critical Infrastructure. The list is limited due to the focus on the specific topic of this report and could be a guideline for the identification and selection of the right ones for the precise Critical Infrastructure (CI) under examination.

In our opinion, one of the most relevant documents is the National Institute of Standards and Technology (NIST) -Artificial Intelligence Risk Management Framework (AI RMF 1.0) [82] which considers the challenges of AI Risk Management (*measurement, tolerance, prioritization, organizational integration and management*). NIST AI RMF explains that AI risk can be defined only in a specific context (AI application). Another relevant point is the definition of the concepts of AI Risks and Trustworthiness as a list of properties of the AI system that characterize a trustworthy AI system (*valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed*). In a real AI system, all these features should be considered and possibly evaluated in conjunction with the context in which the system is going to be deployed.
NIST AI RMF should be considered with some companion documents such as NISTIR 8312[83], NISTIR 8367[84], NIST SP 1270 [85], and the NIST AI RMF Playbook [86]that helps to implement the AI RMF Core functions (Govern, Map, Measure, Manage).

The International Standard Organization (ISO) has a similar approach but with some interesting differences. ISO defined a management standard for deploying and operating AI systems, this document, ISO/IEC 42001:2023 (Information technology — Artificial intelligence — Management system)[87], "*applies the harmonized structure (identical clause numbers, clause titles, text and common terms and core definitions) developed to enhance alignment among management system standards. The AI management system provides requirements specific to managing the issues and risks arising from using AI in an organization*." (see also Chap. 4 of present report)

[82] (National Institute of Standards and Technology (NIST), 2023)
[83] (National Institute of Standards and Technology (NIST), 2021)
[84] (Broniatowski, 2021)
[85] (National Institute of Standards and Technology (NIST), 2022)
[86] (National Institute for Standards and Technology, 2024)
[87] (ISO/IEC 42001, 2023)

There are some standard companion documents like ISO/IEC 22989 (Information technology — Artificial intelligence — Artificial intelligence concepts and terminology)[88] and, specifically for risk management, ISO/IEC 23894:2023 (Information technology — Artificial intelligence — Guidance on risk management)[89] which uses the framework of the well-known ISO 31000[90] and applies it to AI Risk. The approach followed by ISO on AI risk is based on ISO 31000 with specific integration related to the peculiarities of AI use. As an example, ISO 31000:2018, Clause 4 defines several generic principles for risk management. The new standard defines and provides further guidance to Clause 4 on how to apply such principles where necessary: e.g. the AI risk management must be: *Inclusive*, *Dynamic*, based on *Best available information*, consider *Human and cultural factors,* and require *Continual improvement.*
This method of integrating AI into an existing framework is strengthened by the future ISO/IEC CD 27090 (Cybersecurity — Artificial Intelligence — Guidance for addressing security threats and failures in artificial intelligence systems) [91]now (end of January 2025) in Committee Draft. This document will represent the connection between cybersecurity and AI systems that are mainly made by software and, consequently, cybersecurity is a very important issue but, as will be underlined in the following, it is not the most important one.

The European Union Agency for Cybersecurity (ENISA) during the last few years has constantly inserted AI as one of the relevant sources of threats. The Agency has developed many studies on the use of AI in many fields, some of them concerning cybersecurity associated with the development and deployment of Artificial Intelligence systems.
The ENISA report "Multilayer Framework for Good Cybersecurity Practices for AI[92]" provides a layered approach to improve cybersecurity in ICT infrastructure that host components based on artificial intelligence. This approach is centered on risk management:

- at a foundational level according to NIS (EU) directives,
- at the AI level, where typical AI vulnerabilities are discussed along with specific threats, to promote specific security controls and the adoption of appropriate standards,
- at the application context level, where specific threats affecting each sector (energy, health…) are examined.

As it will be stressed in the following this last level of the framework has a prominent importance when it comes to risks related to AI.
The last ENISA Threat Landscape (Sept. 2024) [93]cites AI as an assistant for Threat actors, as a tool for criminals, as a social engineering tool, as a use of chatbots for attacks, and finally as a tool for the defender.

Taxonomies of the various aspects (risk, threat, etc.) of AI, due to the novelty of related technologies and studies, are continuously growing. On the risk taxonomy, there are also some instruments, like an MIT initiative that makes available an AI Risk Repository[94] extracted from scientific papers concerning risks based on two taxonomies: Casual factors and Risk Domain. As an indication of the "chaotic" evolution of AI, it is worth underlining that in this repository over 700 AI risks are categorized by their cause and risk domain.

[88] (ISO/IEC 22989, 2022)
[89] (ISO/IEC 23894, 2023)
[90] (ISO/IEC 31000, 2018)
[91] (ISO/IEC CD 27090, 2024)
[92] (ENISA, 2023)
[93] (ENISA, 2024)
[94] https://airisk.mit.edu/

Another very pragmatic approach is described in the document "Deploying AI Systems Securely. Best Practices for Deploying Secure and Resilient AI Systems"[95] published by nine Cybersecurity Agencies from the USA, Canada, Australia, New Zealand, and Great Britain. The document is synthetic, full of references, and represents a pragmatic and authoritative guide. It is not focused on Critical Infrastructure, but the suggested approach is valid for CI also in consideration that the US CISA (Cybersecurity and Infrastructure Security Agency) is one of the authors.

Specific to the cyber risk derived from the use of AI in Critical Infrastructure there is a MITRE publication "Principles for Reducing AI Cyber Risk in Critical Infrastructure: A Prioritization Approach*"*[96] *based on the risk level of critical infrastructure functions enabled by AI, and the potential for unacceptable outcomes. It is important to define a level of unacceptable consequences before deciding (1) whether to apply AI to a critical infrastructure function and (2), where AI is applied, what lengths to take to ensure its cybersecurity ".* It must be noted that prioritization is mandatory in any risk analysis because it implies funding the most relevant mitigation measures and considering only cybersecurity yields a very partial approach.

The use of AI in Critical Infrastructure and related security issues are addressed by US security agencies like the US Homeland Security in "MITIGATING ARTIFICIAL INTELLIGENCE (AI) RISK: Safety and Security Guidelines for Critical Infrastructure Owners and Operators."[97] The Guidelines are mapped to the NIST AI RMF already cited. It is a synthetic document with a relevant part in the form of a bullet list that "*specifically addresses risks to safety and security which are uniquely consequential to critical infrastructure*."

In the European Union, the most relevant regulation is the EU Artificial Intelligence Act (EU AI ACT) (Regulation (EU) 2024/1689)[98], it is the first law applicable to AI Systems and, despite the criticism raised from the approach of regulating a field still object of research, it is important because it applies to all EU countries and is relevant in defining the danger (risk) of the usage of AI systems. The mentioned regulation establishes a horizontal framework for AI usage and adopts a risk-based approach that categorizes AI applications by potential risk levels defined based on the AI systems usage. Four levels of risk (Low, Medium, High, and Unacceptable) are defined. In addition, this regulation states also a list of prohibited AI practices (article 5) and classification rules for high-risk AI systems (article 6). Under Article 9, "A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems". In this regard, a critical and comprehensive analysis is given also in the paper "Risk Management in the Artificial Intelligence Act"[99].
In Annex III (point 2) the usage of AI systems in CI is classified as High Risk; with some possible exemptions stated in Article 6 point 3.
It must be noted that EU AI ACT does not define Artificial Intelligence but classifies AI systems based on their use irrespective of the type of technology used.
Consequently, AI systems used in CI belong, very likely, to High-risk systems and must be compliant with requirements of section 2 from Art. 8 to Art. 27 and have also notification obligations. One of these requirements is (Art.9) the establishment, implementation, documentation, and maintenance of an adequate Risk management System.

The most relevant point emerging from NIST and ENISA documents is that context is the most relevant aspect to be considered when deciding whether to apply AI, and context is vital in CI

[95] https://media.defense.gov/2024/Apr/15/2003439257/-1/-1/0/CSI-DEPLOYING-AI-SYSTEMS-SECURELY.PDF
[96] (MITRE Christopher Sledjeski, 2023)
[97] (Department of Homeland Security (DHS), 2024)
[98] (European Union, 2024)
[99] https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/risk-management-in-the-artificial-intelligence-act/2E4D5707E65EFB3251A76E288BA74068

because the same AI can have non-relevant or fatal consequences depending on the deployment environment. Definition of the context is not easy, but in general terms and a not exhaustive list can be sketched as the physical environment of the CI (site, machinery, processes, etc.), security and safety requirements, organization and people of the CI, control systems and instrumentation, functions where AI can apply, and type of AI technology to be used.
The rationale for this fact is that every context (and every Critical Infrastructure) has its own hazards to be considered when a risk assessment is conducted. A risk assessment of the AI component not tied to the specific service delivered by the critical infrastructure is not of great use. Having said that, the risk assessment shall consider the contribution to the risk level introduced by the AI components and, possibly, new hazards for the infrastructure imported with the introduction of the AI components.

In cybersecurity, the availability, integrity, and confidentiality of data are particularly relevant, this is because the protection of information is one of the fundamental purposes of cybersecurity itself.

The use of AI systems makes the use of data, often in the form of Big Data collected from many heterogeneous devices, even more relevant. That is, with the introduction of AI systems, data quality becomes even more significant because the quality of responses can vary greatly. In Chapter 1 of this report, the quality of data and its importance were underlined. In the present chapter, it is of interest to see how the use of data also involves risks that must be evaluated.

For example, in a DHS (Department of Homeland Security) paper[100] focused on mitigating the risk of the use of AI in CI, under Risk Category: AI Design and Implementation Failures reference is made to Statistical Bias: *The reproduction or amplification of computational errors or distortions due to data integrity failures or other design defects. This could result in biased outputs and erroneous decision-making.*

It is very important to consider that most AI systems are stochastic in nature and therefore data preparation and normalization are of utmost importance to avoid introducing risks of processing errors.

These brief considerations combined with those in Chapter 1, make clear the importance, often underestimated, of a Data Governance process to adequately assess the risks associated with data quality in conjunction with the type of AI technology adopted. In the absence of such a process, the assessment of the risk derived from "poor quality and security" of data will be very difficult. In the following paragraph 5.4 Data Governance process is briefly explained.

Finally, it is important to mention the current increasing use of complex IRM (Integrated Risk Management) approaches and methodologies by organizations that want to integrate the risks arising from the use of AI tools into the complex of various types of business risks (legal, financial, operational, etc.). However, it is necessary to point out that AI also represents an enabler for the use of more complex Risk Analysis methodologies.

5.2 Risks associated with the AI instrument

We have seen in the previous paragraph the main standards and models for the assessment and management of AI-related risk, in this paragraph we will discuss which risks have been associated with the use of AI as a tool.

[100] (Department of Homeland Security (DHS), 2024)

The improvement of the capabilities of artificial intelligence means that its diffusion increases, and so do the risks linked to potential threats and vulnerabilities of the AI systems adopted by managers of Critical Infrastructure (CI).

The use of AI is recent, and the subject is still in an early stage of application just as many of the tools available today are still being researched and developed. Consequently, the landscape of methodological and technical proposals for assessing the risks resulting from the use of AI as a tool is very diverse.

We believe that one of the most methodologically comprehensive approaches is the ENISA document "AI Cybersecurity Challenges. Threat Landscape for Artificial Intelligence. ".[101]
The document is not very recent (December 2020) but contains two broad taxonomies related to assets and threats.
This document is quite comprehensive because it defines a taxonomy of assets grouped into six categories (*Data, Models, Actors, Processes, Environment /Tools (hardware/ software),Artefacts*).
The corresponding threat taxonomy is grouped into eight typologies (*Nefarious Activity/Abuse, Unintentional Damage, Legal, Failures or malfunctions, Eavesdropping Interception Hijacking, Physical attacks, Outages, Disasters*).
Some of these threats are common to all kinds of IT tools, such as those related to natural disasters, while others are specific to the use of an AI tool, such as model compromise or data poisoning.
The complete threat taxonomy is listed in Fig. 5.1
The ENISA document, through its taxonomies, is useful for gaining an overall view of the complexity that arises from the application of AI tools. However, we aim to provide guidance to help the reader select the best methodological tools for a proper risk assessment of the use of AI to improve resilience in CIs. Proper assessment should be understood as to which is most appropriate to the organizational, operational, and technological context in which the CI operates.

[101] (ENISA, 2020)

Fig 5.1 AI threat taxonomy

*(source**:** ENISA **"***AI Cybersecurity Challenges. Threat Landscape for Artificial Intelligence")*

The use of AI as a tool involves certain risks that are specific to the tool itself. These risks can be divided into two broad categories, one related to attacks on AI systems as such, and a second type related to failures in the design and implementation of an AI system.

Targeted attacks on AI systems supporting critical infrastructure can have a direct or indirect impact on the resilience of critical systems.

The papers cited below are only a part of the literature on the subject and are summarized very briefly. The choice, obviously not exhaustive, is intended to provide an overview of the most significant approaches (in the opinion of the authors) which, as it is evident, are highly differentiated. Differentiation testifies to the current continuing evolution of the subject.

The U.S. Cybersecurity and Infrastructure Security Agency (CISA) [102]has defined a roadmap to AI that is the contribution of this Agency to the national plan led by the DHS and defined in the document "Safety and Security Guidelines for Critical Infrastructure Owners and Operators" [103]already cited. According to CISA and DHS documents, among the attack categories that most frequently impact the resilience of critical infrastructure are:

- Adversarial Manipulation of AI Algorithms or Data: modified algorithms or data may cause the AI systems to behave in unexpected ways.

[102] (Cybersecurity and Infrastructure Security Agency (CISA), 2023)
[103] (Department of Homeland Security (DHS), 2024)

- Malicious injection of prompts into an AI system to bypass the model may cause system malfunction or the disclosure of sensitive information.
- Attacks may render an AI system unavailable to its intended users and interrupt the services.

CISA and DHS documents identify a set of mitigation strategies like:

- Alternate Process Redundancy: Redundant manual operations with physical device operation, traditional computation, and analytics, or other manual tasks that normally benefit from high automation.
- Data Masking: Modifying sensitive data in such a way that it is of little to no value to unauthorized intruders while still being usable by software or authorized personnel.
- Dataset Validation: Efforts to protect the datasets that machine learning and AI algorithms are trained on by filtering poisoned data examples from training, using subject matter expert-annotated datasets, model hardening, two-detector models, and otherwise protecting data from adversary manipulation.

MITRE Company has followed a more detailed approach with a specific Atlas. MITRE is worldwide renowned for its Cybersecurity Atlas which is a recognized reference for assessing and mitigating cyber threats.
MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems)[104], a globally accessible knowledge base of adversary tactics and techniques against Al-enabled systems, lists 60 techniques, such as *Initial Access, ML Model Access Execution, Privilege Escalation, Impact,* with techniques which are specific to AI enabled system – examples are *LLM Prompt Injection, Evade ML model, Denial of ML Service, Erode ML model or dataset integrity*.
The MITRE ATLAS focuses on LLM (Large Language Model) and ML (Machine Learning) which are just some AI technologies, but the ATLAS is an ongoing document and new improvements and updates are continuous.

ENISA in "Securing Machine Learning Algorithms" [105]lists threats and vulnerabilities that specifically apply to machine learning algorithms. This document (issued in 2021) can serve as a basis for the classification of the ML algorithms and their taxonomy. Threats to ML and corresponding vulnerabilities are identified as such related security controls that are mapped, when possible, to ISO or NIST frameworks.

A very pragmatic approach is followed by OWASP in its "Top 10 for Large Language Model Applications" document[106]. OWASP lists the 10 most relevant (in their experience) threats to LLM and for each of them (e.g. Prompt Injection) a list of related vulnerabilities, prevention and mitigation strategies, examples of attack scenarios, reference links, and related frameworks and taxonomies are given.
The OWASP approach is focused on giving operable hints to operators that want to have a first level of protection against the most probable threats and consequently reduce their risks.

AI Design and Implementation Failures.

The CISA and DHS cited documents are also focused on the consequences of AI design and implementation failures. Initial AI lifecycle phases are very important, and due care must be exercised to reduce these kinds of risks that can be synthetically classified as follows.

[104] (MITRE, 2024 on going)
[105] (ENISA, 2021)
[106] (OWASP, 2025)

- Malfunctions or unexpected behavior of AI systems – they may be linked to excessive permissions or poorly defined operational parameters.
- Unintended failure or unexpected behavior of AI systems, when they are confronted with circumstances outside the original range or context.
- Unintentional defects, inconsistent System Maintenance processes, and interoperability with other AI, or not AI-based systems can lead to unexpected or harmful behavior.

A more comprehensive approach to AI lifecycle has been followed by the University of Oxford which developed a tool called *capAI*[107] to support companies in assessing compliance with the EU's Artificial Intelligence Act.
capAi identifies phases and steps of the conformity assessment and describes controls to be checked for each of the phases. These same controls may be used to analyze risks related to an AI-based tool implementation. The evaluation of *capAI* includes control items for all lifecycle (design, development, evaluation, operation, and retirement) of an AI tool.
In the design phase, the organizational governance and the use case should be defined and documented. In the development phase, data and model are documented and checked, and the related risks have been addressed (such as: how to handle missing data or imbalanced data; scaling; and normalization).

5.3 Risks associated with the use of AI

From the previous paragraphs, we have seen that context is relevant for choosing an AI system appropriate to the foreseen usage. Nevertheless, the most appropriate AI system will never be immune to errors or faults and the possible effects of these problems should be evaluated. A possible approach is to define the tasks entrusted to AI systems and their relative level of entrustment. Then it is possible but not easy to identify the possible malfunctions and the possible consequences from which derive a possible evaluation of this type of risk. The more critical is the mission (in the defined context) assigned to the AI systems the more is relevant this phase of risk assessment.

a. Tasks entrusted to AI and level of entrustment
b. Identification of malfunctions and assessment of consequences

Regarding tasks entrusted to AI they can span, for instance, from: predictive analysis, support to decisions, activating scenarios, simulating real environments for training or simulations. Each of these applications must be evaluated in the context of the resilience of a CI.
Hence, we can assume that a specific AI application can be entrusted to implement or support, also separately, single phases of a resilience cycle. With this approach, we can identify specific risks for the resilience of a CI related to failure, malfunction, or compromission of the AI systems.
Threats and vulnerabilities listed in the previous paragraph are helpful as a basis for this identification.
Depending on the level of entrustment to the AI system (from simple support to full autonomy) the associated risk can be assessed and managed.
For example, in the case of a predictive analysis AI-based system, aimed to improve resilience, risks are related to:

- Improper recognition of a threat compromising the prevention function, thus giving way to a following disruption or degradation of service.
- Improper recognition of the degradation level, compromising a timely intervention to activate the restorative phase of the resilience cycle.

[107] (University of Oxford, 2022)

- Activation of a wrong or untimely scenario to restore the performance of the critical infrastructure.

Risks imported through AI systems in Critical Infrastructure are thus reconducted to specific risks affecting the CI itself.

5.4 Tools for mitigation of risks arising from AI

The application of AI systems is relevant when the amount of data to be processed and algorithms complexity exceeds the capabilities of humans. The characteristic of AI systems, different from software programs that operate on algorithms that can be tested in a deterministic way, is that results are based on probabilities or like in neural networks, in a non-re-constructible way. This implies that special attention should be paid to tools, if any, to detect AI malfunctions or possible errors. The use of these tools must also be evaluated as an adjunctive risk (they may have faults and errors) and their possible deployment should be considered when the risk reduction is appreciable in the specified context.

The first means for reduction of risks in using an AI system is surely to adopt a system that either conforms to the requirements of the EU AI Act, if the system is to be used in a European country, or conforms to the requirements, if any, of the country in which it is to be used. We are considering systems in use at CIs and therefore assume, as first approximation, that they are used within the country in which the CI is located. If this assumption is not applicable, compliance with the standards, if any, of the various countries of use should be considered.

For information, as of the date of the last review of this chapter (January 2025), there were more than 1,000 laws and regulations in more than 69 countries or supranational bodies[108].

Compliance with regulations, particularly the EU AI Act, involves complying with several requirements, which in the case of systems to be used in CI will have to conform to the High-Risk level.

The selection of a regulatory-compliant AI system also involves a risk assessment to be carried out according to one of the methodologies outlined in section 5.1, and consequently, controls and technical-organizational measures to mitigate risks will also have to be identified.

It should be remembered that the technologies used for the AI system and the context of usage are fundamental to a proper assessment of the risk in its Safety and Security aspects, giving established compliance with industry standards. However, some general requirements and measures can and should be applied to reduce the risks associated with the operation of the AI system.

These requirements relate primarily to the operational life of the AI system and thus extend beyond regulatory compliance and affect aspects of system upgrading and monitoring that enable the reduction of operational risks and the effects of any incidents by improving recovery capabilities and thus resilience.

An example of these requirements and measures for CI can be found in the DHS document (already cited) [109] that reports the following main usages of AI in CIs as listed by US CISA.

- *Operational Awareness*: This involves using AI to gain a clearer understanding of an organization's operations. For instance, AI can be used to monitor network traffic and identify unusual activity, enhancing cybersecurity.
- *Performance Optimization*: This involves using AI to improve the efficiency and effectiveness of processes or systems. For example, AI can be used to optimize supply chain operations, reduce costs, and improve delivery times.

[108] https://oecd.ai/en/dashboards/overview
[109] (Department of Homeland Security (DHS), 2024)

- *Automation of Operations*: This refers to using AI to automate routine tasks and processes in an organization, such as data entry or report generation. For example, AI can be used to automate the process of sorting and analyzing large amounts of data.
- *Event Detection*: This refers to the use of AI to detect specific events or changes in a system or environment. For example, AI can be used in health monitoring systems to detect abnormal heart rates.
- *Forecasting*: This is the use of AI to predict future trends or events based on current and historical data. For instance, AI can be used to forecast sales trends based on past sales data.
- *Research & Development (R&D*): This refers to the use of AI in the development of new products, services, or technologies. For instance, AI can be used in the pharmaceutical industry to expedite the drug discovery process.
- *Systems Planning*: This refers to the use of AI in the planning and design of systems, such as IT infrastructure. For example, AI can be used to predict the performance of a proposed system under various conditions.
- *Customer Service Automation*: This involves using AI to automate aspects of customer service, such as answering frequently asked questions or processing orders. For example, chatbots are a common application of AI in customer service automation.
- *Modeling & Simulation*: This involves using AI to create models and simulations of real-world scenarios. For example, AI can be used to simulate traffic patterns for urban planning purposes.
- *Physical Security*: This refers to the use of AI in maintaining the physical security of a facility or area. For example, AI can be used in surveillance systems to detect intruders or suspicious activity.

The document warns that these AI use categories are likely to evolve in the future as more complex applications are introduced to Critical Infrastructure.

The use of data (especially Big Data) is fundamental to the operation of AI tools. Consequently, a Data Governance process must be in place as a requirement and tool to reduce the risks of AI use. The topic would require a discussion that is beyond the scope of this report, but given its importance, it is deemed appropriate to mention the key points of a Data Governance process.
Establish a process that:

1. *identifies* the data held by the company according to its use of AI tools
2. *assesses* the possible use, ethically and transparently and otherwise following applicable regulations, of data for AI applications
3. *defines* the operational methods by which legal, ethical, and regulatory compliance of data used during the operational life of AI applications is ensured
4. *ensures* that the data used undergo validation processes, proper preparation (e.g., anonymization, if necessary, normalization, etc.), and adequate protection (protection of integrity, confidentiality, and availability).

Identification of anomalies in AI System

In CIs and particularly for systems that must support mission-critical aspects, duplication, or even triplication of the most important subsystems is often used. Similar considerations apply to instruments for measuring operational parameters, the incorrect evaluation of which can pose a danger to the CI, in this case, alternative parameters and technologies are preferred if possible.

However, while AI systems are often, and more and more widespread, used for detecting anomalies in other systems, it is a very special case when it needs to detect anomalies in an AI system. For this, a different, independent, or more reliable AI system should be used. But this does not appear to be an easy solution. As an example of a possible solution, a specific n-way architecture could be

proposed, with several alternative systems operating in parallel, and the final output is given when there is a concordance between the outputs of a single system, or by a voting mechanism filtering the output of the system affected by an anomaly.

In the ICT sector, at least in malware detection, the simultaneous use of several systems, using different AI technologies or with different parameters, is already applied to reduce the probability of missed detection. A similar criterion could be applied, consistent with economic feasibility, to other areas of the use of AI systems.

This application of AI is also a typical example of AI used to improve defense and resilience mechanisms.

5.5 Risks associated with interdependencies.

AI systems are often used in conjunction with other AI and non-AI systems. Depending on the context, the risks derived from these interdependencies should be assessed.

The consequences of interdependence between systems fall into the case of cascade effects. For this, a holistic approach to risk assessment and management should possibly be adopted. The tools to be used in this case belong to trees or graph representations of interactions between systems (e.g. event tree, fault tree, attack tree). These techniques can be used either when considering the infrastructure under examination as an isolated system, or when consequences on the external environment or other infrastructure are considered.

Generally, we can identify two types of interdependence:

a. AI systems interdependence
b. AI and non-AI systems interdependence

For the aim of this paragraph, the same analytical approach of the previous 5.3 paragraph can be used. When an AI or non-AI (isolated or interconnected) is used, depending on the context, every hazard affecting the effectiveness of that component to the interconnected ones should be identified, and consequential risks evaluated from the perspective of performance or resilience of the system in which they are embedded.

In general, there are no peculiar differences between cases a. and b., having said that non-AI systems are typically characterized by deterministic behavior and a precise function relating to inputs and outputs.

The assessment should be conducted based on the map (tree or graph) of the whole system, considering for each node what are the possible consequences of a fault, a failure, or an error, including the communications channel, and what are the effects on the following node.

5.6 Conclusion

The use of AI technologies in ICs introduces, as usual, both new opportunities and new risks (which according to the ISO definition of risk are two aspects of the same) that can affect the resilience of infrastructure. The landscape within which a risk assessment must be made becomes, if possible, much more complex.

What appears to be fundamental, and this is nothing new in risk analysis, is the absolute relevance of the context under consideration.

There is no general risk analysis, only the application of techniques, increasingly complex with the introduction of AI, that must be applied to the individual CI whose resilience implies extending the analysis to all relationships (functional, organizational, supply, contractual, legal, etc.).

Almost paradoxically, the use of AI technologies is practically indispensable to meet these new challenges.

References

Broniatowski, D. (2021). *NISTIR 8367 Psycological Foundations of Explainabilty and Interepretability in Artificial Intelligence.*
Tratto da https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8367.pdf

Cybersecurity and Infrastructure Security Agency (CISA). (2023). *Roadmap for AI.* Tratto da https://www.cisa.gov/resources-tools/resources/roadmap-ai

Department of Homeland Security (DHS). (2024). *Mitigating artificial intelligence (ai) risk: Safety and Security Guidelines for Critical Infrastructure Owners and Operators.".* Tratto da https://www.dhs.gov/sites/default/files/2024-04/24_0426_dhs_ai-ci-safety-security-guidelines-508c.pdf

ENISA. (2020). *"AI Cybersecurity Challenges. Threat Landscape for Artificial Intelligence".* Tratto da https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges

ENISA. (2021). *Securing Machine Learnig Algorithms.* Tratto da https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms

ENISA. (2023). *multilayer-framework-for-good-cybersecurity-practices-for-ai.* Tratto da https://www.enisa.europa.eu/publications/multilayer-framework-for-good-cybersecurity-practices-for-ai

ENISA. (2024). *ENISA Threat Landscape* .

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence.* Tratto da https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

ISO/IEC 22989. (2022). *ISO/IEC 22989 (Information technology — Artificial intelligence — Artificial intelligence concepts and terminology).*

ISO/IEC 23894. (2023). *IISO/IEC 23894:2023 (Information technology — Artificial intelligence — Guidance on risk management)* .

ISO/IEC 31000. (2018). *ISO 31000:2018 Risk Mnagement -Guidelines.*

ISO/IEC 42001. (2023). *ISO /IEC 42001:2023 Information technology — Artificial intelligence — Management system.*

ISO/IEC CD 27090. (2024). *ISO/IEC CD 27090 (Cybersecurity — Artificial Intelligence — Guidance for addressing security threats and failures in artificial intelligence systems.*

MITRE. (2024 on going). *Adversarial Threat Landscape for AI Systems (ATLAS™).* Tratto da https://atlas.mitre.org

MITRE Christopher Sledjeski. (2023). *Principles for Reducing AI Cyber Risk in Critical Infrastructure: A Prioritization Approach.* Tratto da https://www.mitre.org/news-insights/publication/principles-reducing-ai-cyber-risk-critical-infrastructure-prioritization

National Institute for Standards and Technology. (2024). *NIST AI RMF Playbook.* Tratto da https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook

National Institute of Standards and Technology (NIST). (2021). *NISTIR 8312 Four Principles of Explainable Artificial.* Tratto da https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf

National Institute of Standards and Technology (NIST). (2022). *NIST SP 1270 Towards a Standard for Identifying an Managing Bias in Artificiali Intelligence.* Tratto da https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf

National Institute of Standards and Technology (NIST). (2023). *NIST AI 100-1 Artificial Intelligence Risk Management Framework (AI RMF 1.0).* Tratto da https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

OWASP. (2025). *Top 10 for Large Language Model Applications.* Tratto da https://owasp.org/www-project-top-10-for-large-language-model-applications/

University of Oxford. (2022). *capAI (conformity assessment procedure for AI systems).* Tratto da https://artificialintelligenceact.eu/assessment/

6 A COMMENTARY ON AI APPLICATIONS IN CRITICAL INFRASTRUCTURE *(Alberto Stefanini, Giorgio Pizzi, Glauco Bertocchi, Francesca Della Mea)*

This chapter examines the role of artificial intelligence in three key sectors: energy, transport, and information and communication technology (ICT). Each section explores how artificial intelligence (AI) enhances the resilience, efficiency, and security of critical infrastructure within these domains:

1. **Energy**: AI-driven tools for grid management, fault detection, and renewable energy integration.
2. **Transport**: AI applications in traffic management, predictive maintenance, and autonomous vehicles.
3. **ICT**: AI's contributions to cybersecurity, network optimization, and data center management.

By analyzing specific use cases and identifying cross-sectoral opportunities, this chapter highlights the transformative potential of AI in supporting and safeguarding critical infrastructure.

6.1 AI and Energy Infrastructure Resilience[110] *(Alberto Stefanini)*

As energy demands continue to rise globally, the resilience of energy infrastructure has become increasingly critical. Energy infrastructure includes power generation, transmission, distribution, and storage systems, all of which must withstand a variety of challenges, including extreme weather events, cyber threats, and equipment failures. The integration of innovative technologies, particularly AI, has the potential to enhance the resilience of these infrastructure significantly.

Resilience in Energy Infrastructure

Resilience in energy infrastructure refers to the ability of systems to prepare for, respond to, and recover from disruptive events while maintaining essential functions. This encompasses several aspects, including the physical robustness of infrastructure, the capacity for rapid recovery, and the flexibility to adapt to changing conditions. The increasing frequency of extreme weather events—such as hurricanes, floods, and heatwaves—has highlighted the need for energy systems to not only withstand these challenges but also to anticipate them. For example, the "ClimAdapt" project in Italy aims to assess the impact of climate change on energy infrastructure and develop adaptive strategies utilizing AI for better forecasting and risk assessment[111]. By incorporating climate data and predictive analytics, this project seeks to enhance the resilience of energy systems against the increasing frequency of extreme weather events, thereby ensuring continued service delivery during critical periods.

Resilience Phases: Preventive, Adaptive, and Restorative

[110] This section was generated with the assistance of OpenAI's ChatGPT, an AI language model designed to facilitate and support the drafting process. While every effort has been made to ensure accuracy and relevance, the content has been reviewed and validated by the author to meet the specific objectives of this report. ChatGPT was utilized as a tool to enhance productivity and provide inspiration, and all final decisions and interpretations are those of the author.
[111] (CNR, 2018).

The resilience of energy infrastructure can be categorized into three key phases: preventive, adaptive, and restorative. In the preventive phase, AI technologies enable proactive measures to identify vulnerabilities and mitigate risks before they materialize. For instance, predictive maintenance powered by AI can analyze data from equipment sensors to forecast failures, allowing operators to conduct maintenance activities before an actual breakdown occurs. This capability is exemplified by the ENEL Group's initiatives, where AI is employed to optimize maintenance schedules and enhance the reliability of their grid operations [112]

During the adaptive phase, AI facilitates real-time adjustments to energy management practices based on changing conditions. For example, smart grids equipped with AI algorithms can dynamically balance supply and demand, integrating various energy sources, including renewables, to maintain stability. The Italian project "Smart2Grid" showcases this adaptive capability by utilizing AI to manage the integration of distributed energy resources, thus improving grid flexibility and resilience[113] This approach not only helps in managing fluctuations in renewable energy generation but also enhances overall grid reliability.

In the restorative phase, AI plays a crucial role in recovery following a disruptive event. AI-driven analytics can evaluate the impact of outages and streamline recovery processes by optimizing the allocation of resources and personnel. For instance, post-event analyses using AI can identify areas most affected by power outages, enabling utilities to prioritize restoration efforts effectively. This capability is argued by an authoritative recent article published by Nature[114].

Integrating AI for a Sustainable Energy Future

The integration of AI into energy infrastructure not only enhances resilience but also aligns with broader sustainability goals. The Italian government has set ambitious targets to achieve carbon neutrality by 2050, necessitating significant advancements in energy efficiency and the integration of renewable resources. AI technologies can assist in monitoring emissions, optimizing energy consumption, and promoting sustainable practices among consumers. Initiatives like the "Green Deal, which aims to foster sustainable growth, recognize the importance of AI in driving the transition to a low-carbon economy.[115]

Moreover, AI can play a vital role in promoting energy democracy by empowering consumers to take control of their energy usage. Smart home technologies, integrated with AI, enable consumers to monitor and adjust their energy consumption patterns, leading to more efficient energy use and reduced costs. The "Casa Clima" project in Italy demonstrates how smart home systems can utilize AI to optimize energy consumption and increase awareness of sustainable practices among users[116].

The Role of AI in Enhancing Energy Resilience

In today's complex and interconnected energy landscape, the implementation of artificial intelligence (AI) is emerging as a transformative force, fundamentally reshaping how energy infrastructure operate and respond to challenges. AI's ability to process vast amounts of data, recognize patterns, and make informed predictions plays a crucial role in enhancing the resilience of energy systems, ensuring they can withstand and recover from various disruptions.

AI-Powered Predictive Analytics for Enhanced Resilience

[112] (Best Practice AI, 2024)
[113] (Cabiati , Gianinoni, de Nigris, & Serri, 2021).
[114] (Nearing, et al., 2024)
[115] (Damiani, 2024)
[116] (Klima Haus - Casa Clima, 2022)

One of the most significant contributions of AI to energy resilience is its ability to harness predictive analytics. By analyzing historical data, weather patterns, and operational metrics, AI algorithms can forecast potential failures and disruptions before they occur. For example, AI can predict equipment malfunctions in power plants or transmission lines, allowing operators to perform maintenance proactively. This predictive capability not only minimizes downtime but also optimizes resource allocation and reduces maintenance costs. In Italy, the implementation of predictive maintenance solutions by major energy providers has proven effective in enhancing operational efficiency and reliability.

Real-Time Monitoring and Decision Support Systems

AI enables real-time monitoring of energy systems, allowing for immediate responses to emerging threats or anomalies. Advanced sensor technologies combined with AI analytics provide operators with comprehensive insights into system performance, enabling them to identify issues as they arise. For instance, the "Enel X" initiative employs AI to monitor energy consumption patterns in real time[117], facilitating demand response strategies that adjust supply based on consumption fluctuations. This capability enhances grid stability and resilience by ensuring a balanced supply-demand equation, particularly during peak usage periods.

Moreover, AI-driven decision support systems can aid in emergency response scenarios. By analyzing data from various sources, including weather forecasts, grid status, and consumer behavior, AI can recommend optimal responses to mitigate disruptions. For example, during extreme weather events, AI can help utilities prioritize restoration efforts by predicting which areas are likely to experience outages, thus ensuring a more efficient recovery process.

AI for Cybersecurity in Energy Infrastructure

In an increasingly digital landscape, energy infrastructure face significant cybersecurity threats that can compromise their resilience. AI plays a vital role in enhancing the cybersecurity posture of energy systems by detecting and responding to potential threats in real time[118]. Machine learning algorithms can analyze network traffic patterns to identify anomalies indicative of cyberattacks, allowing for immediate countermeasures. The "CyberSec" initiative in Italy focuses on developing AI-driven cybersecurity frameworks for energy utilities, aiming to protect critical infrastructure from emerging threats[119]. This proactive approach enhances the overall resilience of energy systems by safeguarding against vulnerabilities that could disrupt operations.

Collaboration and Knowledge Sharing for Resilience

The successful integration of AI in enhancing energy resilience necessitates collaboration among various stakeholders, including government agencies, industry players, and research institutions. Collaborative initiatives, such as the "Horizon Europe" program, aim to foster innovation in energy resilience through research and knowledge sharing. By pooling resources and expertise, stakeholders can develop AI solutions that address specific challenges faced by energy infrastructure.

Furthermore, the dissemination of best practices and lessons learned from successful projects is essential for driving widespread adoption of AI technologies. Workshops, conferences, and training programs can facilitate knowledge transfer, ensuring that industry professionals are equipped with the skills needed to implement AI-driven solutions effectively.

[117] (Enel) (Ciurli , 2024)
[118] (Hive Power, 2022) (Schneider Electric, 2024)
[119] (Cyber Security Italia Events, 2023)

Conclusion

In conclusion, the resilience of energy infrastructure is essential for maintaining a reliable and secure energy supply in an increasingly complex and challenging environment. The role of AI in enhancing this resilience cannot be overstated, as it enables proactive measures, real-time adaptations, and effective recovery strategies. By integrating AI technologies into energy systems, Italy can improve operational efficiency, strengthen cybersecurity, and foster a sustainable energy future. As illustrated by various national projects and initiatives, the ongoing collaboration between government entities, research institutions, and industry stakeholders will be crucial in realizing these goals. The path forward involves not only leveraging AI to enhance resilience but also embracing a holistic approach that prioritizes sustainability, innovation, and community engagement in the energy transition.

6.2 Microfactories and Critical Infrastructure Protection: The Case of Valsesia[120] *(Alberto Stefanini)*

As the resilience of energy infrastructure becomes ever more critical in the face of rising global energy demands and multifaceted challenges, innovative solutions at both macro and micro levels are essential. One such approach is the development of microfactories, which exemplify how localized, adaptive strategies can bolster the resilience and sustainability of energy systems.

Microfactories: bridging Innovation and Resilience in Mountain Regions

Microfactories represent a transformative approach to promoting resilience and innovation in mountain areas such as Valsesia, which, besides being a part of the world's second-largest ski area, embodies the ideal environment for pioneering distributed manufacturing.
These facilities, envisioned as **localized FabLabs**, go beyond emergency repairs and play a critical role in fostering a **culture of technological innovation**. Equipped with tools like 3D printers, 3D scanners, and computer-aided design (CAD) software, they enable rapid prototyping and on-demand manufacturing of unique components. This is particularly valuable in remote areas where maintaining large inventories is impractical. By leveraging **vectorized designs**, FabLabs can empower local communities to produce essential parts for repair to local energy systems, or bespoke innovations, contributing to self-sufficiency and economic sustainability.
In terms of **critical infrastructure**, microfactories are invaluable for enhancing the resilience of **electric grids and telecommunications networks**. While modern utility providers, such as ENEL, have minimized outages with robust contingency measures (e.g., helicopter-transported generators), microfactories provide additional value by enabling localized production of smaller-scale, customizable components, such as enclosures, brackets, or repair parts for electrical systems and telecommunications hardware. This reduces logistical delays, strengthens supply chain resilience, and ensures faster recovery from natural disruptions such as snowstorms or landslides.
Moreover, this role is becoming even more essential with the **growing energy demands driven by emerging technologies**, such as **generative AI** and **large language models (LLMs)**. These systems require substantial computational resources, which in turn increase energy consumption and place additional strain on local power grids. By enabling localized manufacturing of energy-related

[120] This section was generated with the assistance of OpenAI's ChatGPT, an AI language model designed to facilitate and support the drafting process. While every effort has been made to ensure accuracy and relevance, the content has been reviewed and validated by the author to meet the specific objectives of this report. ChatGPT was utilized as a tool to enhance productivity and provide inspiration, and all final decisions and interpretations are those of the author.

components and supporting **Edge Computing solutions**, microfactories can help stabilize grids and meet the surging energy requirements efficiently.
In the ICT sector, microfactories are pivotal for producing hardware that supports **Edge Computing**, a distributed computing model that processes data closer to its source. This reduces latency, improves network resilience, and enables real-time applications such as IoT and AI-driven services. As a result, microfactories serve as a foundational enabler of **next-generation digital infrastructure**, ensuring continuity and adaptability even in remote regions.
Finally, the adoption of microfactories as **distributed innovation hubs** aligns with a **franchising model** tailored for alpine valleys. By integrating these facilities into a wider network of FabLabs, local communities gain access to advanced Industry 4.0 technologies, boosting their ability to innovate and thrive in a competitive technological landscape.
Microfactories thus stand at the intersection of **resilience, sustainability, and technological advancement**. They not only reinforce critical infrastructure but also nurture local innovation ecosystems, positioning mountain regions as key contributors to the global shift toward distributed and sustainable development models.

The Role of Artificial Intelligence in Microfactories

Artificial intelligence (AI) plays a central role in enhancing the capabilities of microfactories, particularly in improving resilience and responsiveness. AI-driven technologies enable microfactories to operate efficiently and securely while addressing the dynamic needs of critical infrastructure systems.

1. **Predictive Maintenance**
   Sensors embedded in infrastructure systems, such as electric grids or telecom networks, continuously gather performance data. AI algorithms analyze this data to predict potential component failures, allowing microfactories to preemptively manufacture replacement parts. This proactive approach minimizes the risk of prolonged outages and ensures that infrastructure remains operational during critical moments.

2. **Adaptive Production**
   AI enhances the adaptability of microfactories by dynamically adjusting production priorities based on local demand. During emergencies like natural disasters, AI systems can analyze real-time data to prioritize the production of urgently needed components, such as transformers for energy grids or communication hardware. This ensures optimal resource allocation and faster recovery times.

3. **Cybersecurity**
   AI safeguards microfactory operations against cyber threats, which are increasingly targeting infrastructure systems. By detecting anomalies and potential breaches in real time, AI ensures that locally produced components are secure and reliable. This is particularly critical when manufacturing sensitive components for energy or telecom systems, where security breaches could have catastrophic consequences.

Broader Applications and Scalability: The Future of Microfactories

While Valsesia highlights the localized potential of microfactories, the concept is scalable to other contexts facing similar challenges. Remote regions, urban areas with strained infrastructure, and disaster-prone zones could all benefit from decentralized production.
For regions susceptible to extreme weather, integrating microfactories into disaster preparedness strategies ensures the availability of key components before disruptions occur. Similarly, in regions

grappling with **geopolitical tensions** or global supply chain disruptions, localized manufacturing reduces reliance on external suppliers, enhancing autonomy and resilience.
The **Green Edge Cloud Computing paradigm** further expands the scope of microfactories. Unlike traditional giga-data centers that consume vast amounts of energy with limited efficiency, smaller, decentralized Edge data centers can operate sustainably by leveraging renewable energy sources such as **Mini-Hydro plants**. Located in rural or less urbanized areas, these facilities can integrate agricultural or industrial activities that utilize waste heat, transforming what was once inefficiency into value.
This approach not only aligns with circular economy principles but also addresses the environmental challenges associated with centralized infrastructure systems. As **IoT adoption** accelerates, the need for localized data processing will grow, driving the evolution of microfactories as hybrid production and data centers capable of supporting resilient, sustainable infrastructure systems.

The Path Forward: From Theory to Practice

Although the microfactory concept remains in its early stages, pilot projects like those in Valsesia and forward-looking studies by organizations such as **Hal Service Spa** are paving the way for broader adoption. By exploring the feasibility of decentralized models, these initiatives demonstrate how microfactories can protect vital infrastructure while advancing sustainability goals.
For instance, Hal Service Spa's ongoing **feasibility study** and **White Paper** on sustainable Edge data centers aim to demonstrate how integrating renewable energy sources can transform the current paradigm. With outcomes expected by 2025, this research highlights the practical potential of downsizing data centers to address both logistical and environmental challenges.
As AI technologies advance and the need for resilient infrastructure grows, microfactories are poised to become integral to infrastructure protection. Their **decentralized, adaptive, and responsive** nature offers a transformative solution to some of the most pressing challenges in infrastructure management, from natural disasters to supply chain vulnerabilities. Pilot initiatives serve as blueprints for other regions, illustrating how microfactories can foster a future where critical systems are not only more resilient but also more sustainable and secure.

6.3 Applications of Artificial Intelligence for the Resilience of Transport Systems *(Giorgio Pizzi)*

Regulatory Framework

Within the European Union, Directive (EU) 2022/2557 identifies critical entities across several transport subsectors, including air, rail, waterborne, road transport, as well as Intelligent Transport Systems (ITS) and public transport. In Italy, public transport has also been included under the legislative decree transposing Directive (EU) 2022/2555.
This discussion will primarily focus on public passenger transport, which falls into the category of “critical entities” as defined by Directive (EU) 2022/2557.
The infrastructure related to these entities are considered “critical” to the extent that the services they provide are deemed “essential.” It is the responsibility of critical entities to identify their critical infrastructure, which may include assets, facilities, equipments, networks or systems, or parts of them, which are necessary for the provision of an essential service, as stipulated by Directive (EU) 2022/2557.

Components of a Transport System

According to the definition of critical infrastructure in Directive (EU) 2022/2557 a public land transport system typically includes the following subsystems[121]:

- Structural subsystems:
  - Infrastructure: track, points, level crossings, engineering structures (bridges, tunnels, etc.), rail-related elements of stations (including entrances, platforms, zones of access, service venues and information systems;
  - Energy;
  - Trackside control-command and signaling systems;
  - On-board control-command and signaling systems;
  - Rolling stock.
- Functional subsystems:
  - Traffic management and operations;
  - Maintenance;
  - Telematics applications for passenger and freight services[122].

While metro systems are generally comparable to railway systems, tram systems may have fewer control-command and signaling subsystems, and trolleybus systems do not require tracks.

In cable transport systems, the main subsystems include[123]:

- Cables and cable connections
- Drives and brakes
- Mechanical equipment
- Vehicles
- Electrotechnical devices
- Rescue equipment

and also Infrastructure (route, stations, supports).
Beyond infrastructure in the traditional sense (e.g., civil engineering works like highway or railway viaducts), transport systems also include vehicle fleets (trains, buses) or electrical power supply subsystems for trolleybus networks.
Mobility-as-a-Service (MaaS) platforms enable integrated and multimodal access to the physical transport network through a single channel for information, planning, purchasing, and trip execution. These platforms may fall under the category of Intelligent Transport Systems

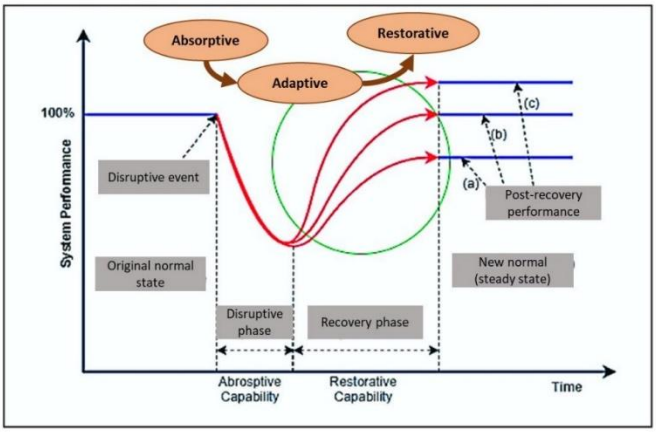
Understanding Resilience

Each of the previously described subsystems is essential for delivering services, thus qualifying them as "critical infrastructure" under Directive (EU) 2022/2557.
The functional subsystems are supported by processes carried out by organizations, meaning that the "elements" in the definition of critical infrastructure are socio-technical, not just physical or technical. The entire network of material, technical, and organizational components under the responsibility of the critical entity can also be considered critical infrastructure.
Resilience is defined as the ability of critical infrastructure (and critical entities) to "prevent, protect against, respond to, resist, mitigate, absorb, accommodate and recover from an incident".
This can be considered an intrinsic quality for a transport system, which is frequently exposed to disruptions - ranging in severity - that can degrade service levels, safety, or lead to complete service disruption.

[121] Directive (EU) 2016/797
[122] Applications for passenger services, including systems which provide passengers with information before and during the journey, reservation and payment systems, luggage management and management of connections between trains and with other modes of transport;
[123] Ref. Regulation (EU) 2016/424

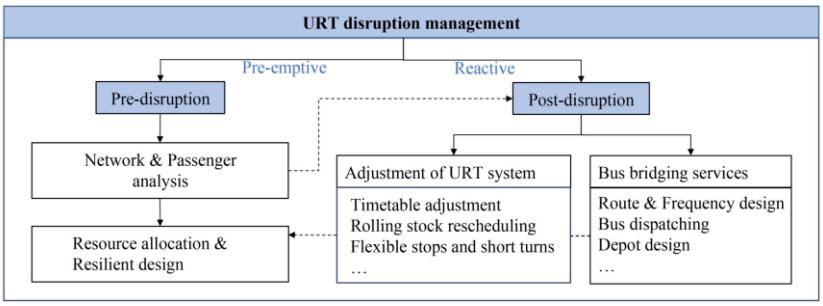
The Resilience Cycle for Transport Infrastructure

The actions involved in preventing, protecting against, responding to, resisting mitigating, absorbing, accommodating and recovering from an incident" in order to restore the operating capacity make up the "resilience cycle."
As already highlighted in other contexts[124], resilience is a multi-faceted concept.
This cycle, or rather this succession of phases, is described by the curve shown in the following figure[125], from which three specific capabilities can be identified (absorb, adapt, restore) that can be fully supported or implemented through artificial intelligence applications.
It is widely and correctly argued that public transport and mobility form the backbone of our society. Therefore, in this context, and according to a view shared with industry operators, the concept of resilience should be applied with an ecosystem approach. As is evident, service interruptions or disruptions to one element cause a degradation that diminishes the service provided to passengers and, in some cases, even compromises their safety.
In transport systems, resilience also addresses operational issues, primarily service regularity and demand satisfaction, not only complete disruptions. These disruptions may be caused by infrastructure failures, track interruptions, equipment or rolling stock malfunctions, cyber-physical attacks, weather events, fires, or sudden spikes in demand.



The management of resilience can be divided into two phases, preventive and reactive, as shown in the following figure[126], which refers to an urban railway system but can, by analogy, be extended to any transport system.

[124] (Bertocchi, et al., 2016)
[125] Immagine da (Sarker & Lester, 2019)adattata da https://ingegneriadellaresilienza.it/la-resilienza-e-i-sistemi-di-gestione/
[126] (Wang, 2024)



Preventive management actions involve:

- The analysis of the network and demand, based on studying passenger behavior and the consequent allocation of resources, i.e., service sizing to achieve the necessary capacity. Based on this, using appropriate modeling, vulnerability analyses can be conducted to optimize capacity, including provisions for certain reserves or redundancy within the transport network's connections.
- The vulnerability and resilience analysis itself, conducted according to the four-phase cycle (preparedness, robustness, recoverability, and adaptability).

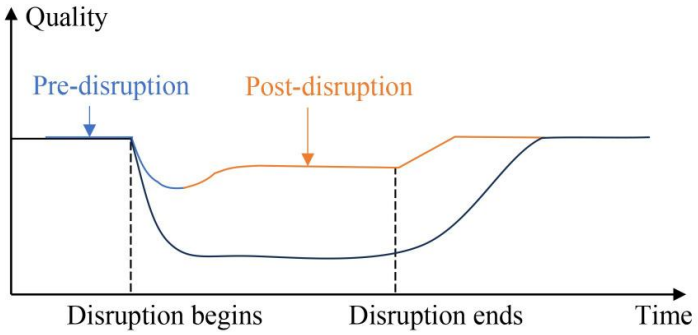
Preventive actions provide the system with "intrinsic resilience" (one could say "by-design"), anticipated in the initial design phase and capable of preventing service degradation and managing the modal shift in demand resulting from interruptions.

Regarding reactive management, the measures consist of:

- Dynamic service rescheduling through modifications to vehicle timetables and frequencies;
- Deploying additional rolling stock to cope with demand spikes caused by external events or local disruptions to the ecosystem;
- Reconfiguring routes in response to track interruptions or infrastructure degradation;
- Providing passengers with accurate information about appropriate behaviours or services to use.

The management of service interruptions shares similarities with managing demand peaks, so the same tools for rescheduling and adaptation can be used.
The management of prolonged interruptions may also involve the implementation of alternative connections using other modes (for example, setting up bus services in the case of disruptions to rail services).
Both regular service planning and emergency management require models that allow for cost minimization while providing optimal sizing criteria, reducing passenger queues and vehicle congestion at stations.
The use of models increases operators' responsiveness in dealing with disruptions and restoring connection capacity.
An additional level of intervention involves implementing on-demand services to allow stranded passengers to move, complementing the capacity provided by bus services. These services create a system integration between vehicles, platforms, and passengers within the area affected by the disruption.
Network analysis tools, vulnerability assessments, and resource allocation and optimization methods used in the design phase, and thus in the preventive approach to resilience, are also employed in the reactive approach, as both service rescheduling and the implementation of additional services require their use.

The purpose of this approach is to reduce the extent or duration of degradation in the event of an adverse occurrence or attack, as illustrated in the following figure[127].



Artificial Intelligence as a Technological Innovation in the Resilience Cycle

If infrastructure or systems are not inherently resilient ("by design"), the challenge is to provide them with this property through external means, in this case, technological.
From a practical standpoint, the key question an artificial intelligence study on public transport should answer is: What problems are solved, and what benefits are expected? In general, for resilience purposes, the benefits include service improvement through threat detection, dynamic and predictive assessment of disruption or degradation risks, and support for response actions based on real-time monitoring and early identification of critical patterns.
Machine learning, predictive analysis, and adaptive responses are the methods through which artificial intelligence can enhance or provide resilience to a transport infrastructure, across various phases of forecasting, assessment, communication, recovery, and recognition of critical conditions.
Resilience, in any case, arises from the interaction between the different components of a system, which is why it can be achieved through monitoring and reconfiguration — activities in which the performance and effectiveness of artificial intelligence orchestrating their interaction play a specific role.

Artificial Intelligence Applications in the Resilience Cycle

A first classification of the various artificial intelligence applications to the phases of the resilience cycle is based on the timeframe in which they are effective (short, medium, long term). This classification places them in the phase of reaction and recovery, adaptation and reconfiguration, or planning and absorption of external disruptions. In the latter case, we could also refer to "by-design" resilience applications.
Apart from this, the concept of resilience by definition leads to a preference for applications that act in the short and medium term, providing reactivity to the system or transport operator, which is a fundamental requirement at the technical-organizational level during the adaptation and recovery phase of service levels. This requirement can also be seen from the passenger's perspective, as the user of the service is the primary individual to protect.
From the literature review[128], it appears that artificial intelligence applications have been evaluated for the following purposes.

[127] (Wang, 2024)
[128] (Jevinger, Zhao, Persson, & Davidsson, 2024)

In the short term:

- Estimating travel and arrival times;Improving communication with passengers;
- Traffic monitoring;Improving dispatching;
- Estimating the number of passengers on vehicles (passenger counting);
- Route selection;Video surveillance for security purposes;
- Supporting diagnostics, event prediction, and maintenance;
- Monitoring passenger status/comfort;
- Supporting emergency management;
- Providing recommendations to travelers;

In the medium term:

- Supporting timetable redefinition;
- Analyzing traveler behavior;
- Supporting service replanning.

These applications fall within the area of operational support and the improvement/adaptation of service levels, and they are aimed at both operators and travelers.

In the long term:

- Studying traveler behavior and analyzing their opinions;
- Analyzing the spatial distribution of users;
- Analyzing the overall state of the transportation ecosystem.

These applications primarily pertain to the areas of service planning, demand analysis, and territorial analysis.

Let’s examine now the specific artificial intelligence applications within each phase (or capability) of the resilience cycle: preventive, absorptive, adaptive, restorative.

Preventive

- Scenario analysis and recommendations for passengers, operators, and planners
- Recognition of situations that could lead to delays, with decision support tools for traffic control and automated problem resolution
- Conducting stress tests, forecasting, and analyzing failure or disruption times based on prepared models
- Providing information to support organizational resilience
- Simulating incidents and disruptions using digital twins, where threats can be implemented to analyze necessary resources and size emergency services
- Simulating the effectiveness of various timetables and schedules.

Absorptive

- Vehicle occupancy assessment
- Decision support.

Adaptive

- Travel and route replanning
- Adapting supply to demand

- Service rescheduling
- Managing demand by shifting to alternative modes.

Restorative

- Decision support
- Service rescheduling.

It is recognized by industry operators that integrating artificial intelligence into digital twins of transport systems is useful for training organizations to handle unforeseen events and to test the implementation of procedures to be applied during the recovery and restoration of operational capacities.

6.4 AI application in the ICT Sector: an outline *(Glauco Bertocchi, Francesca Della Mea)*

The ICT sector forms the backbone of modern digital infrastructure, encompassing data centers, cloud services, and enterprise networks.

This sector was among the first to use the various technologies classified as AI, in the early 2000s, to defend against so-called malware and its continuous and multifaceted evolutions. Therefore, antivirus tools can recognize the behavior of suspected malware and thus acquire the ability to "recognize" even malicious software never previously encountered and not classified as such. Of course, these tools, like all technologies based on fundamentally stochastic processes, also have a certain failure rate and false positives, but at present they are present in every computer system.
At the same time, the application of AI technologies has extended to other areas of the ICT world, particularly the protection of networks, which require constant monitoring that must analyze thousands of messages per second and to the detection of advanced malware such as APTs (Advanced Persistent Threats).
The development of the cloud and the consequent diffusion of very powerful and complex data centers, both software and hardware, has further supported the development of systems with AI capabilities for operational management (load balancing, configuration management, etc.) and predictive maintenance These tools have benefited from the deep learning techniques developed in the last decade and subsequently the diffusion and availability of LLM applications is enabling many companies to use specialized chatbots to support SIEM (Security Information and Event Management) and SOC (Security Operation Centre) operators. These latter tools, SIEM and SOC, also make use of AI techniques to recognize possible adverse or critical events and correlate them with other events that have occurred or are occurring.

No specific application for digital infrastructure resilience appears to be available yet, but there are many products that use AI techniques to improve one or more of the factors that make an infrastructure more resilient. Remember, as mentioned in the introductory chapter that resilience, like security, is a process that has technical, organizational and human components as the people who must define, govern and implement the process.

Particularly since Governance is one of the cornerstones of current standards for AI risk management and beyond (see Chap. 3), it is very likely that in the future this function will be performed by AI tools that will progressively have a management role that will also affect the entire resilience cycle.

The following is a very brief and certainly not exhaustive illustration of the various areas of the ICT sector in which AI tools are currently being used. The mention of products or their manufacturers is intended to indicate some existing tools; it should not be intended as exhaustive, let alone imply any judgment or endorsement on the part of the authors. Moreover, the rapid evolution of the field will soon make such indications obsolete.

ICT (Information and Communication Technology)	**Network Security**: Utilizing AI to detect and mitigate cybersecurity threats in real-time.
	End Point Protection Using AI to detect and mitigate threats coming from end points.
	Cloud Data Protection Use of AI to protect data in cloud ensuring their availability, integrity and confidentiality
	Data Center Optimization: AI-driven management of data centers to improve energy efficiency, load balancing, and resource allocation.
	Predictive Maintenance: Using AI to forecast IT infrastructure (hardware and software) maintenance, helping to reduce shut-off time and disruption of service.
	Asset management. Using AI for inventories and management of configurations of hardware and software components
	SIEM and SOC management. Using AI to manage, classify and report the events registered through Siem and Soc.
	Automated IT Support: Deploying AI-based virtual assistants and chatbots for real-time IT support and troubleshooting.
	Software development and Verification and Validation Building and deploying correct and secure software is a very complex task in which AI tools can reduce the time to market without decreasing quality and security

Network Security

There are many products available on the market that use AI technologies for threat detection and mitigation. Among the vendors we can point to Darktrace, founded in 2013, which uses machine learning to monitor overall network behavior and identify threats. Other vendors include Cisco, SparkCognition, Fortinet, Microsoft, and many others that use AI-containing components to identify new threats.

End Point Protection

Protection of End Points, i.e., machines used by end users to access information systems is one of the key capabilities to reduce threats. In fact, end points are the preferred gateway for many types of attacks, and the detection of suspicious or simply “dangerous” behavior in the context in which the end user operates is vital to prevent or mitigate the propagation of an attack toward the heart of the information system. The use of AI technologies is necessary to ensure adequate resistance to attacks that may occur in unpredictable ways. Manufacturers of tools for this function include IBM, Dark Trace, Cylance, and many others.

Cloud Data Protection

Data protection in the cloud is critical to ensure, among other things, also the capability for ready and complete restart, i.e., enabling recovery with data intact and available. Recovery is a key component of a system's resilience. The use of tools using AI technologies is very important to reduce recovery time, especially in complex systems that require dynamic management according to context. Tool vendors include Darktrace, IBM, Forcepoint, and others.

Data Center Optimization

Areas in which the use of AI techniques can be useful are: energy efficiency by, for example, optimizing cooling according to workload and ambient thermal conditions; workload balancing of servers and other components in order to reduce latency times during peaks; predictive maintenance that is based on the actual utilization of components and their behavior schedules outages in order to minimize them.
All major cloud service providers also offer Data Center Optimization tools. They include Google[129], Microsoft, IBM, Intel and others

Predictive Maintenance

Using AI to forecast IT infrastructure (hardware and software) maintenance, helping to reduce shut-off time and disruption of service. The use of tools with AI technologies enables forecasting based on the analysis of Big Data originating from ICT systems[130]. This functionality is often included in asset management tools. Among the vendors IBM, C3.ai, and other

Asset management

Asset management is an area of application for tools that include AI to perform classification and aggregation functions by purpose, technology, operational requirements, etc. Among the vendors: IBM, Orangelogic, Clarifai, and others

SIEM and SOC management

These tools are among the main beneficiaries of AI technologies for operational management of information systems. Event classification and recognition of “dangerous” events along with correlation with other events that have occurred or are in progress is a typical task where the use of AI can be an essential support for threat mitigation and recovery, i.e., a valuable contribution to resilience.

Automated IT Support

The use of specialized chatbot tools can be of great help to security practitioners at all stages in which a threat is recognized, corrective or mitigation measures are identified, any damage is limited, and finally a return to full operations is worked on. Producers of these tools include Microsoft, Open AI, ChatBot, and many others.

Software Development and Verification and Validation

The use of automated tools in software development is also related to the development methodologies adopted by the company (e.g. DEVPOS) and the technical peculiarities of the software environment adopted. These are specialized professional tools that have widespread use within companies that produce software for their own systems or for others. The use of AI components (e.g., AI assistant)

[129] (Google AI Blog, 2020)
[130] (Gartner, 2022)

within these tools is widespread although not always explicitly stated. Producers include Tricentis, Jetbrains, Zebrunner, Sahi and others

Conclusion

The application of AI tools in ICT is pervasive and the areas mentioned are indicative and not exhaustive. When considering that AI tools are software systems it appears consequent that ICT systems are naturally a field of application for AI. The fast evolution and complexity of the ICT sector make necessary the usage of tools with AI technologies to "govern" them effectively and efficiently. Effectiveness and efficiency make systems more secure and more resilient but with the limitations of the AI tools used.

References

Bertocchi, G., Bologna, S., Carducci, G., Carrozzi, L., Cavallini, S., Lazari, A., Traballesi, A. (2016). *Guidelines for Critical Infrastructure Resilience Evaluation.*

Best Practice AI. (2024, Nov. 11). *AI Case Study - Enel is reducing operational and capital expenses by predicting maintenance and improving asset performance using machine learning.* Retrieved from Best Practice AI: https://www.bestpractice.ai/ai-case-study-best-practice/enel_is_reducing_operational_and_capital_expenses_by_predicting_maintenance_and_improving_asset_performance_using_machine_learning

Cabiati, M., Gianinoni, I. M., de Nigris, M., & Serri, L. (2021). *Sintesi delle attività sulle Smart Grid e il sistema energetico, a supporto delle istituzioni in ambito nazionale e internazionale.* Milano: RSE - Ricerca sul Sistema Elettrico. Retrieved Nov. 11, 2024, from https://www.rse-web.it/wp-content/uploads/2022/11/21009193.pdf

Ciurli , S. (2024, July 11). *Enel's innovation through AI.* Tratto da Enel Group: https://www.enel.com/media/word-from/news/2024/07/ai-future-on-the-road-to-innovation

CNR. (2018, 01 12). *L'intelligenza artificiale esplora il clima e trova conferme e novità - PRESS RELEASE.* Tratto da Consiglio Nazionale delle Ricerche: https://www.cnr.it/en/press-release/7875/l-intelligenza-artificiale-esplora-il-clima-e-trova-conferme-e-novita

Cyber Security Italia Events. (2023). *La Cybersecurity nell'era dell'AI.* Retrieved Nov. 27, 2024, from https://www.cybersecitalia.events/presentazione/

Damiani, E. (2024, May 29). *Intelligenza artificiale: un alleato chiave per il Green Deal europeo.* Retrieved Nov. 27, 2024, from Network Digital 360: https://www.agendadigitale.eu/smart-city/intelligenza-artificiale-un-alleato-chiave-per-il-green-deal-europeo/

Enel. (n.d.). *The advantages of Enel X Energy Management.* Retrieved Nov. 27, 2024, from Enel Energia for the Free Market: https://www.enel.it/en

Gartner. (2022). *The Predictive Power of AI in IT Infrastructure Management.*

Google AI Blog. (2020). *DeepMind AI Reduces Google's Data Center Cooling Costs.*

Hive Power. (2022, Sept. 22). *Intelligenza artificiale e apprendimento automatico nella distribuzione di energia.* Retrieved from Hive Power: https://www.hivepower.tech/it/blog/artificial-intelligence-and-machine-learning-in-energy-

distribution#:~:text=L%27AI%2FML%20%C3%A8%20la%20tecnologia%20che%20gestisce%20le%20reti,i%20dati%20degli%20utenti%20per%20prendere%20decisioni%20accurate.

Jevinger, Å., Zhao, C., Persson, J. A., & Davidsson, P. (2024). Artificial intelligence for improving public transport: a mapping study. *Public Transport, 16*, 99-158.

Klima Haus - Casa Clima. (2022). *Agenzia per l'Energia Alto Adige - CasaClima.* Tratto da Klima Haus: https://www.agenziacasaclima.it/it/chi-siamo-5.html

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., . . . Yossi, M. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*(627), 559-563. Retrieved Nov. 27, 2024, from https://www.nature.com/articles/s41586-024-07145-1

Sarker, P., & Lester, H. (2019). Post-Disaster Recovery Associations of Power Systems Dependent Critical Infrastructure. *Infrastructure*.

Schneider Electric. (2024, Nov. 11). *Le nostre soluzioni basate sull'intelligenza artificiale.* Retrieved from Schneider Electric: https://www.se.com/it/it/work/solutions/artificial-intelligence/solutions.jsp

Wang, L. J. (2024). Urban rail transit disruption management: Research progress and future directions. *Frontiers of Engineering Management, 11*, Frontiers of Engineering Management.

7 ETHICAL AND SOCIETAL IMPLICATIONS *(Luigi Carrozzi, Raffaella D'Alessandro)*

Artificial Intelligence applications may raise important ethical and societal issues impacting individuals, groups and society at large. This chapter introduces such issues identifying possible approaches to manage the negative impacts of AI. Furthermore, a special focus is devoted to the ethical and environmental impacts of AI applications for the Resilience of Critical Infrastructure

7.1 AI and Critical Infrastructure resilience: ethical and societal concerns *(Luigi Carrozzi)*

Artificial intelligence is increasingly used in the management of critical infrastructure, particularly for project development, security, maintenance and performance optimization. But Artificial Intelligence applications may give rise to important ethical concerns with significant negative impacts on human being. Autonomy and self-determination, discriminations and respect of human rights may be endangered by AI outputs. These consequences are potentially brought by any applications of AI that may directly or indirectly impacts on individuals, group of individuals and the society at large. Among others, are worth to be mentioned:

- privacy violations, since the massive amount of data collected may involve processing of personal data violating rights of individuals to privacy and protection of personal data,
- impacts on workforce, since may replace certain job roles, causing unemployment,
- biased output, since the training data may mis-represent or propagate errors relating certain aspect of real word, for example about demographics, leading to possible discriminations,
- security, since AI systems are susceptible to malicious attacks to data and models, seriously compromising the systems' outputs,
- explainability, compromising the right of subjects impacted by AI decisions to understand the inner system's logic that led to that decision,
- misinformation and "deepfakes", that may have serious impacts on individuals and society by influencing opinions and social behaviors,
- violations of intellectual property since generative AI outputs may pose a risk to rightsholders of intellectual property,
- environmental impacts, due to the increasing demand for energy and water required by IA server farms.

Given the above, any AI application that supports operations and resilience of Critical Infrastructure should be passed under scrutiny to evaluate such potential impacts.

European Commission through its High Artificial Intelligence Level Group (AI HLEG) has provided a valuable guiding document "*Ethics Guidelines For Trustworthy AI*"[131] providing direction to adopt ethic principles in AI systems. The value of the document consists in the fact that those ethical

[131] European Commission – Indipendent High-Level Expert Group on Artificial Intelligence set up by European Commission (2019) *Ethics Guidelines For Trustworthy AI.*

principles are derived from fundamental rights established at EU level and whose respect is legally binding in EU. The document, in citing those rights, states in particular that:

"I'ts Among the comprehensive set of indivisible rights set out in international human rights law, the EU Treaties and the EU Charter, the below families of fundamental rights are particularly apt to cover AI systems." And furtherly:

"Many of these rights are, in specified circumstances, legally enforceable in the EU so that compliance with their terms is legally obligatory. But even after compliance with legally enforceable fundamental rights has been achieved, ethical reflection can help us understand how the development, deployment and use of AI systems may implicate fundamental rights and their underlying values, and can help provide more fine-grained guidance when seeking to identify what we should do rather than what we (currently) can do with technology"

The guidelines, after identified the following principles,

- **Respect for human autonomy,**
- **Prevention of harm,**
- **Fairness,**
- **Explicability,**

provide a guidance on the implementation of Trustworthy AI, adopting seven requirements (although identified as non-exhaustive, to be considered apt to include systemic, individual and societal aspects) that should be met, built on those four principles. Methods (technical and non-technical) are then made available for the implementation of those requirements throughout the AI system's life cycle.

The seven requirements are the following:

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, minimization and reporting of negative impact, trade-offs and redress)

And it's worth of note that the AI HLEG Guidelines are specifically mentioned in recital 7 of AI ACT[132] as follows:

"In order to ensure a consistent and high level of protection of public interests as regards health, safety and fundamental rights, common rules for high-risk AI systems should be established. Those rules should be consistent with the Charter, non-discriminatory and in line with the Union's international trade commitments. They should also take into account the European Declaration on

[132] Regulation (EU) 2024/1689

Digital Rights and Principles for the Digital Decade and the Ethics guidelines for trustworthy AI of the High-Level Expert Group on Artificial Intelligence (AI HLEG)"

Notably, the AI ACT in Annex III(2) includes Critical Infrastructure within High-risk AI systems, as referred to in Article 6(2), as follows:

"*Critical infrastructure: AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity*" .

Relating to privacy legislation the General Data Protection Regulation[133] at art. 35 (and article 27 of Directive *2016/689* [134]) provides that a "*Data Protection Impact Assessment* "(DPIA) is mandatory when the processing of personal data is "*likely to result in a high risk to the rights and freedoms of natural persons*"; moreover art. 27 of the AI ACT provides a "*Fundamental Rights Impact Assessment*" (FRIA) for High-Risk AI system. The relations between DPIA and FRIA in the AI ACT are addressed to in art. 27(4), that states:

"*If any of the obligations laid down in this Article is already met through the data protection impact assessment conducted pursuant to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, the fundamental rights impact assessment referred to in paragraph 1 of this Article shall complement that data protection impact assessment*".

7.2 Managing ethical and societal impacts of AI solutions *(Luigi Carrozzi)*

To properly face such AI's ethical and societal impact, it's necessary to be aware of the characteristics of the AI solution and the related ethical and societal concerns that its adoption and use may give rise. These concerns should be considered in the overall AI risk management process, performing a specific ethical and societal impacts assessment to enable a clear picture of the individual and the collective value to protect. Only a sound awareness of the value at the stake may support the proper identification of measures apt to prevent possible harm to people and society at large.

For this objective the adoption of an overall management framework may be helpful. The ISO 42001[135] standard provides a robust framework for the adoption of an AI management system.

According to ISO this standard is "*for organizations of any size involved in developing, providing, or using AI-based products or services. It is applicable across all industries and relevant for public sector agencies as well as companies or non-profits*".

Considering the specific, highly innovative and powerful characteristics of AI technologies, this standard may provide an important guideline to fully exploit the benefits of AI while reducing the negative impacts, including the ethical ones.

Generally speaking, the management of an AI system necessarily should start identifying the full picture of the nature, components of the systems and the contest in which is going to be adopted. The data acquisition process including the provenance, their quality and their preparation process is then

[133] Regulation (EU) 2016/679

[134] Known as "LEA - Law Enforcement Directive": DIRECTIVE (EU) 2016/680 *on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.*

[135] ISO/IEC 42001. (2023). *Information technology — Artificial intelligence — Management system*

crucial to be fully aware of what is in input to the AI system. As the same, the characteristics of AI inference engine adopted and, where applicable, the related decision-making power granted to the system. This may be considered the starting point of a knowledge creation and management process that an organization needs to put in place to have a sound control of IA technologies.

Top management should define an internal organization for AI, assigning AI roles and responsibilities, defining internal policies enabling, among others, the alignment of adoption and use of AI artifacts with the overall intent and the business processes and procedures of the organization.

Moreover, every management system needs to be fed with the necessary resources. Among others, AI requires specific competences and personnel that is able to combine the business culture of the organization with the awareness of the potentials and risks of the technology adopted in each specific business sector.

An AI system should be considered a continuous evolving system that should be necessarily managed along its lifecycle adopting the necessary procedures and precautions as well as a suitable documentation of what the AI systems performs when in use, what are the components and its outputs are also essential for the necessary information to the third parties concerned.

The success of the third party's relationship with business partners, suppliers, customers and civil society is crucial to really take advantage of all the benefits of AI.

In this regard, is the case to mention annex B of the ISO 42001 standard that provides the following implementation guidance for AI controls.

- **Policies related to AI,** that includes: AI policy**,** Alignment with other organizational policies and Review of the AI policy.
- **Internal organization,** that includes: AI roles and responsibilities and Reporting of concerns.
- **Resources for AI systems,** that includes: Resource documentation, Data resources, Tooling resources, System and computing resources and Human resources.
- **Assessing impacts of AI systems** that includes: AI system impact assessment process, Documentation of AI system impact assessments, Assessing AI system impact on individuals or groups of individuals, Assessing societal impacts of AI systems.
- **AI system life cycle,** that includes: Management guidance for AI system development and AI system life cycle**.**
- **Data for AI systems,** that includes: Data for development and enhancement of AI system**,** Acquisition of data**,** Quality of data for AI systems**,** Data provenance and Data preparation.
- **Information for interested parties,** that includes**:** System documentation and information for users**,** External reporting**,** Communication of incidents and Information for interested parties.
- **Use of AI systems,** that includes: Processes for responsible use of AI systems**,** Objectives for responsible use of AI system and Intended use of the AI system.
- **Third-party and customer relationships,** that includes: Allocating responsibilities**,** Suppliers and Customers

7.3 Main ethical and environmental impacts using AI for the Resilience of Critical Infrastructure *(Raffaella D'Alessandro)*

AI technology can pose large-scale risks to humanity, including acute harms to individuals, large-scale harms to society, environmental impact and even human extinction[136]. These risk types include AI impacts that are bigger than expected, worse than expected, willfully accepted side effects of other goals, or intentional weaponization by criminals or states.

In this paragraph we focus on the ethical and environmental risk that can arise using the following AI application for the Resilience of Critical Infrastructure: Predictive Maintenance, Risk Management and Cybersecurity.

More specific details on Impacts and Risk of ethics and environmental issues using AI for the Resilience of Critical Infrastructure are described in the Appendix: "The use of AI application in Critical Infrastructure for Resilience and related risk and impacts on ethics and environmental issues".

Predictive Maintenance

AI can provide operators with enhanced, earlier warnings of potential equipment degradation or failure. This could allow operators to prioritize the equipment most in need of maintenance, improving reliability and preventing costly failures before they occur. AI-based predictive maintenance has already been deployed to support many Critical Infrastructure, from energy, wind turbines to oil and natural gas compressors/pumps, and offers great potential for battery electric storage systems and distribution transformers. AI based predictive maintenance is also used to monitor and prevent equipment failure in non-energy sector like: Wastewater Treatment Systems, Transport, Railways, Ports and Maritime Transport, Airports and Civil Aviation, Telecommunications Networks, Radio and Television Transmission Systems.

Main Ethical and Environmental Impacts and Risks using AI Predictive Maintenance applications

The use of AI for Predictive Maintenance, depending on the specific data processed in each Critical Infrastructure, can lead to the following impacts and risk of ethical and environmental issues:

- data privacy concerns related to extensive monitoring of equipment of physical persons,
- potential job displacement,
- environmental impact of producing AI hardware and disposing of old equipment and outdated infrastructure,
- environmental impact due to energy consumption of AI systems.

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. This information can be used to prioritize investments in resilience measures and develop specific contingency and evacuation plans. In the Financial sector AI models assess also financial risks and optimize investment strategies, for example, predicting market trends.

Main Ethical and Environmental Impacts and Risks using AI Risk Management applications

[136] Andrew Critch, Stuart Russell. (2023). *TASRA: a Taxonomy and Analysis of Societal-Scale Risks from AI*. arXiv:2306.06924, https://arxiv.org/abs/2306.06924

The use of AI for Risk Management, depending on the specific data processed in each Critical Infrastructure, can lead to the following impacts and risk of ethical and environmental issues:

- ethical concerns include the potential for biased decision-making and data privacy,
- environmental risks involve the energy consumption of AI systems.

Cybersecurity

Artificial Intelligence brings many benefits, but disruption of AI could, in the future, generate impacts on scales and in ways not previously imagined. These impacts, at a societal level and in the context of critical infrastructure, include disruptions to National Critical Functions.[137]

AI cybersecurity systems detect and respond to cyber threats in real-time, protecting the integrity of the ICT and AI assets and infrastructure. For example, identifying and mitigating DDoS attacks but also AI specific attacks as for example poisoning attacks, Evasion attacks and Data extraction Attacks[138].

One promising application is in enabling machine-speed analysis of operational technology (OT) and information technology (IT) data, in an effort to identify, detect, and mitigate cyber intrusions into infrastructure before they cause an impact.

Main Ethical and Environmental Impacts and Risks using AI Cybersecurity applications

The use of AI for Cybersecurity, depending on the specific Cybersecurity functions performed, can lead to the following impacts and risk of ethical and environmental issues:

- data privacy concerns related, for example, to extensive monitoring of user account activities,
- ethical concerns, that include the accuracy and transparency of AI decisions,
- environmental risks involve the energy consumption of AI systems.

Furthermore, as AI application in Critical Infrastructure evolves, also ethical and environmental impacts and risks do. So, it is important to monitor these impacts and risks and a very useful resource is the MIT AI Risk Repository where we can find very interesting and updated specific AI Risk related also to ethical and environmental impacts, they are constantly updated by MIT. The MIT AI Risk Repository[139] is very useful to identify and understand updated specific risk related not also to ethics and environmental impacts. Critical Infrastructure can use the MIT AI RISK REPOSITORY to:

- Conduct internal risk assessments,
- Identify new, previously undocumented, risks,
- Evaluating risk exposure and developing risk mitigation strategies,
- Develop research and training.

In Appendix, are provided examples of AI application adopted in main Critical Infrastructure in order to enhance resilience and presented the related risk and impacts on ethics and the environment.

[137] (https://www.mitre.org/news-insights/publication/principles-reducing-ai-cyber-risk-critical-infrastructure-prioritization#:~:text=This%20paper%20identifies%20and%20builds%20on%20existing%20best,decide%20to%20regulate%20AI%20use%20in%20critical%20infrastructure)

[138](https://www.enisa.europa.eu/sites/default/files/publications/Cybersecurity%20and%20privacy%20in%20AI%20-%20Forecasting%20demand%20on%20electricity%20grids.pdf)

[139](https://airisk.mit.edu/#:~:text=The%20Causal%20Taxonomy%20of%20AI,False%20or%20misleading%20information%E2%80%9D)

References

Regulation (EU) 2024/1689 - Artificial Intelligence Act

Regulation (EU) 2016/679 - General Data Protection Regulation

DIRECTIVE (EU) 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

ISO/IEC 42001. (2023). Information technology — Artificial intelligence — Management system

Andrew Critch, Stuart Russell. (2023). TASRA: a Taxonomy and Analysis of Societal-Scale Risks from AI. arXiv:2306.06924, https://arxiv.org/abs/2306.06924

8 CASE STUDIES *(Silvano Bari, Sandro Bologna, Giorgio Pizzi)*

This chapter examines case studies demonstrating AI's impact on the resilience of sectors like healthcare, transportation, water distribution and electricity networks, highlighting its role in creating more robust and adaptable systems.

These sectors are essential to societal function and their resilience is critical for both public well-being and economic stability: Healthcare systems are vital for maintaining public health, especially in emergencies, as they provide essential services like diagnosis, treatment, and emergency care. Transportation networks ensure the movement of goods and people, which is key for economic activities and societal functioning. Water distribution is essential for both daily life and sanitation while electricity powers everything from homes to industries, supporting daily operations and technological advancements. If any of these infrastructure fail or are disrupted, the consequences can be severe, affecting everything from individual lives to national economies. This is why ensuring their resilience, particularly through innovations like AI, is so crucial.

8.1 Applications of Artificial Intelligence for the Resilience of Healthcare Environment *(Silvano Bari)*

Artificial Intelligence (AI) can play a critical role in integrating hospital systems and ensuring facility resilience. Here are some ways AI can help:

Data management optimization: AI can analyze and manage large volumes of data from different systems (such as electronic health records, safety management systems, and building management) to ensure a consistent and timely flow of information.

Workload prediction: using machine learning algorithms, hospitals can predict peaks in activity (e.g., based on historical data) and plan resources such as staff and equipment accordingly, improving operational efficiency.

Safety monitoring: AI can complement surveillance and safety management systems to detect anomalies or suspicious behavior in real time, increasing safety for patients and staff.

Telemedicine and remote care: AI tools can support telemedicine, enabling remote diagnosis and monitoring, facilitating access to healthcare services and reducing pressure on hospital facilities.

Process automation: AI can automate repetitive tasks, such as patient registration or inventory management, freeing up staff for more important tasks and better engaging with patients.

Predictive health analytics: Through the analysis of clinical data, AI can detect critical health conditions early, enabling timely interventions and improving clinical outcomes.

Systems integration: AI can facilitate the integration of disparate systems, ensuring that information is accessible and usable consistently across different units and departments.

Education and decision support: AI-based tools can support clinicians in decision-making by providing data-based recommendations and best practices, improving the quality of care.

By implementing these solutions, a hospital can not only improve operational efficiency but also ensure a more resilient and responsive response to healthcare challenges.

Let’s now explore how artificial intelligence technologies are transforming healthcare efficiency, patient care and operations, by examining success stories of artificial intelligence applications in healthcare.
There are several case studies where artificial intelligence has helped healthcare facilities recover from disruptions or difficulties in hospital services, often through advanced data management, predictive analytics, and process automation tools. One notable example is the use of AI during the COVID-19 pandemic, when many healthcare facilities faced an overload in services.
An example of AI implementation in hospitals is the use of artificial intelligence to predict the evolution of hospitalizations and the demand for intensive care beds. This approach has been particularly valuable during the COVID-19 pandemic, when hospitals faced overwhelming numbers of patients and the need to efficiently allocate resources such as intensive care (ICU) beds.
In this case, AI models analyze large volumes of data, including patient demographics, medical history, current health status, and real-time admission rates. By identifying patterns and trends, these models can forecast the number of patients expected to require hospitalization or intensive care. Hospitals use these predictions to better manage their capacity, ensuring that sufficient resources (e.g., ICU beds, ventilators, medical staff) are available when needed most.
AI-driven prediction tools help hospitals optimize their response to sudden surges in patient numbers, enabling more effective planning and decision-making. This not only improves patient outcomes by ensuring timely care but also helps to reduce the strain on medical staff and hospital infrastructure during peak times.

Case Study: Cleveland Clinic and AI for Patient Management

Cleveland Clinic, a major healthcare system in the United States, sees multiple opportunities for AI to improve operational efficiency, drug discovery, and drug discovery.

In the area of operational efficiency, the Cleveland Clinic is applying AI to run its business smarter and more efficiently, using real-time data to better predict bed availability, patient admissions, staffing levels, and waiting times to be able to handle a greater influx of patients.
In this way, Cleveland Clinic now has an accurate forecast of the number of patients who will enter the hospital in the next 24 hours and can create a better plan for staffing, achieving a 7% increase in daily admissions and providing better service to patients who receive necessary care faster.

AI also allows healthcare providers to spend less time entering and retrieving data from systems and more time caring for patients. The Cleveland Clinic is experimenting with ambient listening software. This AI technology can listen to patient appointments and generate notes for the doctor directly in the medical record. As the clinician continues to review notes, the tool saves clinicians valuable time compared to the traditional data entry required after each patient visit.

Cleveland Clinic has used AI to address disruptions and overloads in services caused by the pandemic. During the peak of COVID-19, they had to reorganize resources, address staffing shortages, and ensure patients received appropriate care despite the huge influx of cases.
Here's how AI helped:

Resource Management and Workflow Optimization: An AI system was implemented to monitor bed occupancy and predict future demand for hospital resources (beds, ventilators, staff). This allowed them to optimize resource allocation in real time.

Telemedicine and Virtual Triage: Using AI, the Cleveland Clinic created a chatbot-based virtual triage system that allowed patients to report symptoms and receive an initial assessment online. This reduced the flow of people in the "emergency rooms" and allowed resources to be reserved for the most urgent cases.

Diagnosis and predictive treatment: Using machine learning algorithms, the hospital was able to identify patients most at risk of complications from COVID-19 early and customize treatment plans accordingly. AI also helped with electronic medical records management, improving the speed with which data was processed.

Decision-Making Support: AI has supported doctors in choosing the most effective treatments by analyzing large amounts of clinical data to identify patterns and predict the outcome of therapies. This has helped improve medical decisions and reduce response times.

By introducing these tools, Cleveland Clinic has been able to maintain a high level of care despite the challenges of interrupting standard services and increasing demand[140].

Case Study: Children's Hospital of Pittsburgh

The Children's Hospital of Pittsburgh of UPMC (University of Pittsburgh Medical Center) in the USA is known for its pioneering use of technology to enhance patient care, safety, and operational efficiency.

This hospital has partnered with Johnson Controls to implement an integrated technology system, including a converged IP network that supports various systems like security, fire alarms, nurse calls, and real-time patient tracking. The hospital has set new standards in pediatric care and digital hospital design, making it one of the most advanced children's hospitals in the US

Challenge: The goal was to build a new, seamlessly integrated children's hospital that prioritizes patient care, safety, and operational efficiency while minimizing human error and managing costs effectively.

Project Overview:

The hospital, known for its exceptional clinical services and pediatric care standards, experiences over a million visits annually, including 13,500 inpatient stays and 258,000 surgeries. There was a critical need for a secure, future-proof, and cost-effective converged network that could be easily scaled and utilized.

The hospital adopted a comprehensive integration of clinical, business, and building systems to facilitate the creation of a fully connected infrastructure that enhanced patient care, improved safety, and optimized operational costs across all facilities.

The automation of the Children's Hospital of Pittsburgh did involve advanced technologies, including elements of artificial intelligence (AI). The implemented systems were designed to create smart, data-driven infrastructure that helped automate several critical processes, such as patient flow

[140]https://www.forbes.com/sites/randybean/2024/10/06/how-cleveland-clinic-is-innovating-in-healthcare-with-data-analytics-and-ai/
https://my.clevelandclinic.org/-/scassets/files/org/giving/newsletter/using-artificial-intelligence-to-beat-covid-19
https://health.clevelandclinic.org/ai-in-healthcare
https://www.smithsonianmag.com/science-nature/how-doctors-are-using-artificial-intelligence-battle-covid-19-180977124/

management, energy efficiency, and security systems. AI tools were used to analyze data from these systems in real-time, making predictions and adjustments to optimize operations and improve outcomes for both patients and staff.

For example, predictive algorithms helped enhance patient care by anticipating maintenance needs for medical equipment, optimizing energy use based on hospital occupancy, and improving safety protocols through intelligent monitoring systems. These kinds of AI applications are increasingly common in hospital automation to ensure operational efficiency and enhanced decision-making.

Other examples of AI uses include:

Advanced diagnoses: AI is used to analyze medical images, such as X-rays and MRIs, to detect conditions such as tumors or heart abnormalities. This helps doctors get faster and more accurate diagnoses.

Personalization of treatments: AI is also used to analyze patients' genetic and clinical data, allowing doctors to design personalized treatment plans more precisely.

Predictive care: AI technologies can continuously monitor patients, analyze data in real time, and predict possible complications, such as infections or cardiac arrests, allowing for timely interventions.

Robotics: The use of AI-assisted robots in surgery is increasing. Robots can assist surgeons during complex operations, improving accuracy and reducing recovery time for patients.

These examples are part of a broader effort to use AI to improve medical outcomes and hospital efficiency[141].

In Italy too, AI has demonstrated that it can offer crucial support not only in managing emergencies, as in the case of the pandemic, but also in improving long-term sustainability and in dealing with the pressure on healthcare systems[142].

Case Study: Humanitas Research Hospital in Milano-Rozzano

Humanitas Hospital in Milan used AI to address pandemic challenges and improve patient management[143].

Some key points of the AI application were:

[141] https://www.johnsoncontrols.com.au/-/media/jci/insights/2015/be/files/be_cs_childrens_hosptial_pittsburgh_upmc.pdf

[142] IRCCS Istituto Clinico Humanitas was ranked 1st among Italian hospitals, and 34th out of 250 worldwide in the World's Best Smart Hospitals 2021 ranking by Newsweek. Evaluation criteria included the deployment of the most advanced technologies, the use of Artificial Intelligence, robotic surgery, telemedicine and the presence of digital services. Besides Humanitas, 13 other Italian hospitals made it in the prestigious international ranking: Ospedale Pediatrico Bambino Gesù di Roma, San Camillo-Forlanini di Roma, Ospedale San Raffaele, Policlinico Universitario Agostino Gemelli, Casa Sollievo della Sofferenza di San Giovanni Rotondo, Istituto Giannina Gaslini, Policlinico Campus Bio-medico di Roma, Ospedali Riuniti Marche Nord, Ospedale Niguarda di Milano, Centro Cardiologico Monzino, Istituto Europeo di Oncologia, Meyer-Azienda Ospedaliera Universitaria di Firenze.

[143] https://www.hunimed.eu/news/newsweeks-worlds-best-smart-hospitals-2021-humanitas-ranked-first-italian-hospital/

Automated Triage and Telemedicine: Humanitas used AI algorithms to develop triage systems that allowed them to efficiently manage patients remotely. Through telemedicine platforms and chatbots, patients could report symptoms, receive preliminary consultations, and obtain indications for treatment at home or to access hospital care only if necessary. This helped reduce overcrowding in hospital facilities, maintaining the efficiency of essential services.

Clinical data analysis and clinical outcome prediction: The hospital implemented machine learning systems to analyze patients' clinical data, identify risk factors, and predict the evolution of the clinical conditions of COVID-19 patients. This allowed doctors to intervene more quickly and in a personalized way, improving treatment outcomes.

Resource Optimization: During the most critical phases of the pandemic, Humanitas used AI to monitor the availability of beds, ventilators and other critical resources, predicting demand and ensuring that the most serious patients had access to intensive care. AI also helped manage medical staff shifts more efficiently, reducing the stress and fatigue accumulated by healthcare workers.

Support for medical research: Humanitas has partnered with technology companies to use AI in clinical research, especially to analyze large volumes of genetic and molecular data to better understand the virus and its effects. This has accelerated the process of discovering and validating new treatments and therapeutic protocols.

Artificial Intelligence for research and treatment: Humanitas has established an AI Center that combines data analysis and machine learning with the hospital's clinical and research efforts. The goal is to enhance personalized patient care, improve the accuracy of treatments, assist in diagnoses, and optimize patient flow, leading to overall improvements in healthcare and hospital management. The center doesn't just process clinical data, but also develops intelligent algorithms that identify patterns, find correlations, and create predictive models to drive innovation in Predictive Medicine and Imaging Diagnostics.

Case Study: Policlinico Agostino Gemelli in Rome

Ospedale Policlinico Universitario Agostino Gemelli, one of Italy's leading medical centers, has launched several AI-based initiatives to improve operational efficiency and quality of care. During the COVID-19 pandemic, the hospital adopted several technological solutions to address the disruption of hospital services and improve patient management[144].

AI for COVID-19 Patient Management: Gemelli used machine learning algorithms to analyze COVID-19 patient data and predict the course of the disease. This analysis allowed them to identify patients at risk of developing more severe forms of the disease, allowing for timely interventions. In addition, AI was used to predict the evolution of hospitalizations and the demand for intensive care beds, thus optimizing resource management.

Triage and telemedicine: Like other facilities in Italy, Gemelli has developed AI-supported telemedicine systems to remotely monitor patients with mild symptoms or recovering from COVID. This has reduced the pressure on hospital facilities and allowed doctors to provide continuous care to patients, without overloading emergency departments.

Oncology Research Project: Gemelli collaborated with several technology companies to develop AI applications in the field of oncology. The project used AI to analyze large volumes of clinical and

[144] https://www.policlinicogemelli.it/news-eventi/covid-19-al-gemelli-intelligenza-artificiale-e-big-data-aiutano-a-fronteggiare-la-pandemia/
https://gemelligenerator.it/it/facilities/real-world-data/

genetic data to improve early cancer diagnosis and personalize treatments. This was especially important during periods of service disruption, when resources were limited and cancer patients needed special attention.

Robotics and Automation: The hospital has also used AI in robotics to improve the efficiency of surgical operations, with AI-assisted robotic systems enabling minimally invasive procedures, reducing risk to patients and speeding up recovery times. Again, the pandemic has accelerated the adoption of such technologies to improve operational efficiency.

8.2 Applications of Artificial Intelligence for the Resilience of Transport Systems *(Giorgio Pizzi)*

Use Cases - Real-world examples of AI applications contributing to transport system resilience

The PRECINCT Project

The PRECINCT Project (Preparedness and Resilience Enforcement for Critical INfrastructure Cascading Cyberphysical Threats and effects with a focus on district or regional protection)[145] has developed, through living labs, case studies to improve the resilience of interdependent critical infrastructure in the event of a cyber-physical attack. Among the critical infrastructure considered in the project are transport systems. One of the project's goals is to create collaborative infrastructure for managing cyber-physical resilience among the various structures, allowing critical entities to establish AI-enabled ecosystems and enhanced resilience support services.

In the PRECINCT project, artificial intelligence is used to provide critical infrastructure operators with timely information, derived from data, that supports decision-making processes. These include predictive maintenance, what-if scenarios, incident anticipation, and mitigation planning. Identifying situations that help critical infrastructure operators restore them to optimal conditions also leads to a better return on investment.

The project also considered public transport systems as a "reserve for resilience" to be used during disruptions to primary infrastructure services, providing alternative routes and services to passengers, as well as useful solutions for critical infrastructure operators. Additionally, public transport plays a crucial role in maintaining the regularity of evacuation operations. Its strategic integration with critical infrastructure contributes to overall resilience in urban systems.

One of the living labs was developed in Ljubljana and focused on collaboration between the national railway network, urban bus transport, the electricity distribution operator, and the telecommunications infrastructure, with a connection to the local police.
The Ljubljana living lab concentrates on preventing the effects of cyber-physical attacks on transport and mobility hubs, also introducing the management of simultaneous DDoS attacks targeting the electricity and telecommunications operators.

The Athens living lab of the Precinct project involves three critical infrastructure operators providing primarily transport services across the wider area. Enhancing resilience and rapid coordination in the

[145] https://www.precinct.info/

event of unforeseen incidents among critical infrastructure operators yields enormous benefits. In crises induced by natural disasters or intentional attacks, communication between the affected parties can drastically reduce negative impacts and cascading effects, leading to quicker recovery and ensuring safe evacuation during emergencies. This also benefits the economic development of the area.

The Precinct project highlights the importance of a systemic approach to resilience.
AI-based Technologies Available or in Use by Public Transport Operators and Their Application for Resilience
Direct discussions with experts in the mobility and public transport sectors also highlight that resilience is a systemic requirement and must be achieved at both the technical and organizational levels.
Technology, in this case artificial intelligence, is expected to support the related processes.
Although there is no known existence of a complete "suite" for managing or supporting resilience within a transport infrastructure, it can be noted that individual phases of the resilience cycle can be supported by AI through applications or products that are already available. In some cases, these are already in use by transport operators, though the extent of their adoption varies.
Below, we will mention the main implementations of artificial intelligence, including a broader experience carried out by a major railway operator.

Video Analysis for Passenger Counting

This application, already in use by several public transport operators, is based on an artificial neural network trained for image recognition to provide a reliable count of the number of passengers on board, as well as those boarding and alighting at stops. This data is then used to build and update origin/destination matrices, which are essential for service planning. In general, the issue of passenger counting and monitoring vehicle overcrowding proved to be of great importance during the pandemic, when legal regulations imposed a maximum limit on the number of passengers on vehicles. In effect, this was the implementation of restoring operational capacities in response to an unforeseen event (the spread of the virus) that hindered the safe use of the mobility system. This application can be placed in the adaptive or restorative phase of the resilience cycle.

Video Analysis for Parking Monitoring

This application supports a real-time information service on the availability of free parking spaces in a specific area. This application is linked to resilience, particularly in the preventive phase, as it helps prevent congestion—significantly degrading the service—caused by vehicles circulating while searching for parking.

Urban Mobility Applications

Artificial intelligence is used to predict changes in air pollution levels and provide decision support for traffic restrictions, modulating them and helping to avoid drastic reductions that would result in disruptions to vehicle usage. This application also falls within the predictive phase.
These applications, based on machine learning and predictive analysis, enable the optimization of both public and private transport.

Passenger Information Applications

AI-based chatbots are increasingly being adopted by transport operators to provide specific guidance to passengers. These are particularly useful in cases of service disruption or degradation, ensuring

that passengers interactively receive the necessary information to properly utilize mobility services and meet their needs.

This application relates to the adaptive and restorative phases of the resilience cycle.

Attack Prevention

In the preventive phase of the resilience cycle, artificial intelligence applications are aimed at recognizing conditions that indicate potential attacks.
The Prevent-PCP project[146] uses artificial intelligence to detect unattended objects, preventing the consequences of attacks on transport infrastructure (e.g., railway stations, subways, or intermodal hubs).
The Shield4Crowd project[147] focuses on identifying vulnerabilities in public spaces, such as stations and intermodal hubs, and provides solutions for crowd control, which are particularly important during evacuation phases.

Predictive Maintenance

The availability of elevators is particularly important in subway and railway stations. Predictive maintenance based on machine learning allows interruptions and maintenance activities to be planned, preventing unplanned disruptions and service degradation.
The same technique is used in the rail transport sector to prevent failures in the track infrastructure (e.g., switches) and vehicles, scheduling maintenance activities to avoid unexpected service interruptions.

Service Monitoring and Optimization

Some providers offer platforms that use artificial intelligence for short, medium and long-term forecasts to support in service planning and provide decision support tools for managing disruptions. In particular, vehicle and driver allocation, along with service and schedule planning, are supported by artificial intelligence.
However, according to information gathered from industry operators, these platforms have not yet reached a sufficient level of maturity for effective use in the field.

Disruption Prediction

A transport system is constantly exposed to disruptive events of varying impact and, therefore, must continuously adapt its configuration to maintain consistent service levels. In doing so, it must constantly implement the resilience cycle.
The Swiss Federal Railways (SBB) manage a high-density, highly interconnected rail network. With the increasing demand for transportation, it became necessary to expand capacity and the ability to prevent or mitigate the impact of unforeseen events, as disruptions in such a tightly interconnected network could affect an entire region.
As a result, machine learning was adopted to quickly recognize conditions that lead to service degradation (disruptions of various magnitudes) and reduce their impact on both the infrastructure and rolling stock.

[146] https://prevent-pcp.eu/
[147] https://shield4crowd.eu/

The machine learning solution is applied to a data lake containing information on past disruptions, the causes of service failures, and weather forecasts[148], resulting in a system that continuously improves its performance.
The system is capable of predicting disruptive events in advance and preventing their effects by implementing the necessary actions to maintain an acceptable service level, reacting to interruptions, and dynamically rescheduling services.
The methodology adopted is based on the reinforcement learning[149]

8.3 Applications of Artificial Intelligence for the Resilience of Electric Infrastructure *(Sandro Bologna)*

Case Study: AI assistant supporting human operators' decision-making in managing power grid congestion

The AI use case presented here following has been inspired from the EU Project AI4REALNET[150] , and is based on the Call from RTE Région ile de France "*Entreprises et chercheurs paticipez au Challenge AI pour la Transition énergétique"*[151] , April 2023. The description of the application is reported in the Document "*Description Challenge RTE*" that can be downloaded from the Call web page.
The AI assistant oversees the transmission grid, using SCADA (Supervisory Control And Data Acquisition) data and available EMS (Energy Management System) tools to identify issues and categorize them for human intervention. It monitors power flow, voltage, and balance, adhering to defined operational conditions. Anticipating problems, it sends binary alerts to the operator with confidence levels, avoiding excessive alerts to maintain operator focus (i.e., controls attention budget). Action recommendations include topological changes, storage adjustments, redispatching, and renewable energy curtailment. The human operator selects an action or seeks more information, exploring alternatives. After the operator's decision, the AI assistant provides feedback through load flow calculations, logging decisions for continuous learning and interaction improvement.

Different modes of interaction are possible between AI assistant and human operator, ranging from "full human control" to "full AI control". The selected mode depends on the industry domain and context. In this use case, an ex-ante choice is made to apply a hybrid interaction where the human operator gets the final word on AI assistant recommendations.

Here is the following list of different steps covered by the AI assistant:

The AI assistant monitors the situation of the transmission grid by using the available data from SCADA and EMS tools and categorizes issues by distinguishing the ones needing intervention by the human operator.
The situation of the transmission grid is monitored at the appropriate horizon (e.g., a few hours ahead to 30 minutes ahead) by using relevant forecasts (generation, consumption). Issues correspond to deviations from acceptable operation conditions of the electric system, mainly defined by:

[148] https://www.zuehlke.com/en/case-studies/sbb-delivers-smooth-operations-with-an-intelligent-early-warning-system
[149] https://www.netcetera.com/stories/news/20200109-sbb-flatland-challenge-ai.html
[150] EU Project AI4REALNET https://ai4realnet.eu/
[151] https://www.iledefrance.fr/toutes-les-actualites/entreprises-et-chercheurs-participez-au-challenge-ia-pour-la-transition-energetique

- Power flow on electric lines not exceeding thermal limits (considering, for instance, a tolerance for temporary overload).
- Voltage maintained within a defined range.
- Generation and load are always balanced (frequency is maintained around 50 Hz).

The AI assistant monitors these operating conditions and considers a predefined list of contingencies according to the operational policies of the TSO (Transmission System Operator), which include:

- The nominal grid, i.e., the "N" situation (in which all grid elements are available).
- Cases in N situations where overload duration exceeds allowed thresholds: depending on TSO's operational policies, it can be indeed allowed to let transit flows exceed a temporary threshold on a given line (e.g., flows can be higher than x A for 20 minutes, after which line will automatically trip). Note: such equipment is used on all lines of RTE's grid (Réseau de Transport d'Electricité).
- A list of possible "N-1" (electric system's state after the loss of one grid element and possibly several grid elements depending on the TSO's policy).

When anticipating issues requiring intervention, the AI assistant raises alerts for decisions at the appropriate horizon (e.g., a few hours ahead down to 30 minutes ahead) to the human operator in time to carry out corresponding actions. These alerts are "binary" in the sense that either the AI assistant sends a persistent alert or not, and they are associated with a level of confidence, i.e., the level of certainty of the AI assistant that the electric system won't remain within acceptable operation conditions if no action is performed. The level of confidence is based on the uncertainty in the forecasts. The AI assistant should not send too many alerts to keep the human operator concentrated on his or her tasks and thus ease his or her workload.

For a given alert, the human operator receives action recommendations from the AI assistant, with information on the predicted effect and reasons for the decision. Possible actions are:

- Topological action: topology can be changed by switching power lines on and off or reconfiguring the busbar connection within substations.
- Redispatching action: change the flexibility's (generator, load, battery, etc.) active setpoint value. Redispatching actions include therefore storage actions (e.g., define the setpoint for charging and discharging storage units such as batteries)
- Renewable energy curtailment: limits the power output of a given generation unit to a threshold, defined, for example, as the ratio of maximal production Pmax (a value of 0.5 limits the production of this generator to 50% of its Pmax).

The human operator chooses a proposed recommendation or requests new information or explanations or looks for a different action guided by an exploration agent or via manual simulation using other specific tools (that aren't part of the AI assistant).

The human operator performs the needed actions according to his/her decision. The AI assistant provides feedback to the human operators on the corresponding effects: this is performed afterward (1 hour or more after the facts) by running a load flow calculation.

The decisions made are logged with their corresponding context to continuously learn from realized actions and improve the interactions between the human operator and the AI assistant (e.g., relevance of proposed recommendations for actions).

This use case only addresses congestion issues, even if other types of issues can arise on the Transmission Grid and are handled by the operators (e.g., voltage).

8.4 Artificial Intelligence for Water Networks *(Silvano Bari)*

The Italian water network loses one billion cubic meters of water per year: it is estimated that over 40 percent of all the drinking water introduced into the country's distribution network is wasted due to leaks. These losses constitute a major criticality, exacerbated by climate change and the lack of summer rainfall, which generate situations of extreme variability between periods of heavy rainfall and periods of drought.

In addition to updating infrastructure, therefore, it is essential to improve the detection and repair of leaks, to avoid significant waste and supply problems for the population.

In recent years, research in the sector has made enormous strides: artificial intelligence techniques, together with the implementation of intelligent sensors, can offer a significant contribution to the automatic monitoring of water losses, helping to make water networks more resilient and sustainable and reducing waste. Furthermore, machine learning models, based on the huge amount of data available from *smart meters*, i.e. intelligent meters that allow precise measurements at a distance, allow overcoming traditional methods of finding leaks, such as visual inspections, analysis of acoustic signals and vibrations[152].

So, such a system of automatic monitoring can use a combination of sensors and machine learning to detect leaks or anomalies in a water distribution system.
This system can use flow and pressure sensors to continuously monitor the water network. When an anomaly occurs, the data is analyzed using a deep learning algorithm trained to recognize abnormal patterns. This enables the system to detect leaks or other problems in real-time, triggering alerts to operators and allowing for quick intervention.

Here's a general idea of how such a system works:

1. Sensors for data collection

The system relies on flow and pressure sensors that are installed throughout the water system. These sensors measure key parameters like:

- water flow: the rate at which water is flowing through pipes;
- water pressure: the force exerted by the water in the pipes.

The sensors continuously send real-time data to a central monitoring system, where it can be analyzed for any irregularities.

2. Data preprocessing

The raw data from the sensors is likely cleaned and pre-processed before analysis. This might involve:

- removing noise or irrelevant information;
- normalizing the data to account for different flow rates or pressure conditions in various parts of the system.

[152] https://serviziarete.it/wp-content/uploads/2023/09/luglio-agosto-2023-Lintelligenza-artificiale-per-le-reti-idriche.pdf

3. Anomaly detection using deep learning

A deep learning algorithm (such as a neural network) is trained to recognize normal behavior patterns in the system. For instance:

- when water flows through pipes without leaks, the flow and pressure should behave within certain expected ranges;
- if a leak occurs, it can cause pressure drops or irregular flow rates in certain sections of the pipe network.

The deep learning model is trained on historical data, which includes normal conditions as well as data from past leaks or anomalies. Once trained, the model can identify:

- leaks: changes in flow or pressure that are inconsistent with normal conditions, such as sudden drops in pressure or spikes in flow;
- other anomalies: blockages, system inefficiencies, or equipment failures.

4. Alert system

When the deep learning model detects an anomaly, it triggers an alert. The system may:

- send notifications to maintenance personnel or operators;
- highlight the location and severity of the anomaly on a map of the water network;
- provide actionable insights, like the possible cause of the issue (e.g., leak, blockage, etc.).

A system like this described could also incorporate *feedback loops* where the model improves over time as more data becomes available. This allows the system to become more accurate at detecting leaks or anomalies, especially in different seasons or under varying usage patterns.

Some additional considerations:

- Edge computing: Sometimes, processing is done directly at the location of the sensors to reduce latency and bandwidth usage.
- Integration with maintenance systems: the anomaly detection system may be connected to a larger asset management system, allowing maintenance teams to track and prioritize repairs.

Here are some real examples of application.

The **Laboratory of Thermofluid Dynamics at the Free University of Bolzano** has developed a system that, through the use of flow and pressure sensors located in the water system, is able to report any type of leak or anomaly and highlight it, through a deep learning algorithm.

The system, developed thanks to the growing availability of data (*big data*) collected by smart meters in water networks, is based on *graph neural networks (GNNs)*, mathematical models of neural networks enhanced by the integration of graph structures. This enables a spatial understanding of water networks, fundamental for the identification of anomalies. The data relating to pressure and flow in the system's pipes are used to more quickly and accurately detect anomalous events that could signal a failure or a leak[153].

The **HERA Group**, an Italian multi-utility that operates over 35,000 km of water distribution network, has adopted tools based on artificial intelligence to guide the maintenance activities of its aqueduct network and in an effort to reduce the level of water leaks, through a complex assessment of the risk of pipe breakage and its consequences.

The approach, developed in collaboration with ISOIL Industria S.p.A and Rezatec, combines satellite data with artificial intelligence techniques to produce risk maps expressing parameters such as the *Likelihood of Failure (LOF)*. The algorithm takes into account not only the diameter, historical data

[153] Zanfei A., Menapace A., Brentan, B.M., Righetti M., Herrera M. "Novel approach for burst detection in water distribution systems based on graph neural networks" in https://www.sciencedirect.com/science/article/abs/pii/S2210670722004073

series on breakages or the age of the pipes, but also the level of the aquifers, the type of soil and the ambient temperature[154].

Engineering has developed a digital solution for detecting water losses through the acquisition and digitalization of information on a *GIS* platform, which supports hydraulic engineering experts in managing their activities.
Hydraulic network modeling enables preliminary analysis of collected data and topographic surveys, while leak pre-localization occurs using simulation tools based on artificial intelligence and data collected by sensors, with information acquired in real time to evaluate and manage critical issues on a day-to-day basis[155].

A research project carried out by **Terranova company with the University of Florence** focuses on the use of Artificial Intelligence to improve the management of water networks, studying and developing innovative techniques for monitoring and identifying leaks.
The objective consists in the detection of water leaks through *Anomaly Detection* models capable of prefiguring an ideal situation (for example, a pressure map without leaks) and comparing it with real data from the network. This approach allows identifying unexpected trends due, for example, to leaks, failures of the water network or pumping systems, anomalous consumption, all through the analysis of data collected in real time.
The research includes the use of smart sensors and data analysis techniques based on artificial intelligence[156].

[154] https://www.isoil.it/categoria-news/un-progetto-italiano-usare-lintelligenza-artificiale-per-efficientare-le-reti-idriche/ https://www.rezatec.com/solutions/water-utilities/pipeline-risk/
[155] https://www.industriaitaliana.it/engineering-infrastrutture-idriche-ia/
[156] https://www.terranovasoftware.eu/news/intelligenza-artificiale-terranova-innova-il-servizio-idrico https://www.trilance.com/news/unifi-la-borsa-di-dottorato-finanziata-da-terranova-per-l-individuazione-delle-perdite-nelle-reti-idriche

9 CONCLUSIONS

In examining the intersection of Critical Infrastructure Resilience and Artificial Intelligence, this report highlights the potential and challenges associated with integrating AI into systems vital for societal well-being. Through the exploration of resilience, ethical and societal implications, and risk management, several key conclusions emerge:

1. **Enhancing Resilience with AI**: AI technologies can significantly enhance the resilience of critical infrastructure by improving real-time monitoring, predictive analytics, and automated decision-making. These capabilities enable infrastructure systems to anticipate, resist, and recover from disruptions more effectively.

2. **Historical Perspective on AI in Critical Infrastructure:** AI's integration into critical infrastructure has evolved from experimental applications to essential components in sectors like energy, transportation, and healthcare. Historical advancements illustrate AI's potential but also highlight the challenges of adoption, such as dependency risks, technical failures, and the need for robust supervision to avoid unintentional failure. Lessons from past implementations underscore the necessity for careful, incremental integration of AI to mitigate these risks effectively.

3. **Ethical and Societal Implications**: The deployment of AI in critical infrastructure brings ethical and societal considerations to the forefront. Issues such as algorithmic bias, transparency, and accountability must be addressed to foster public trust. The adoption of AI must also account for its impact on employment, social equity, and privacy, ensuring that the benefits of technological advancements are distributed fairly.

4. **Risk Management and AI Integration**: AI introduces new dimensions to risk management, offering tools to identify vulnerabilities and mitigate risks proactively. However, these systems also create novel risks, including cybersecurity threats, dependence on AI-driven processes, and the potential for adversarial manipulation. Effective integration of AI requires a multi-layered approach to risk management, combining technological, organizational, and policy measures.

5. **Balancing Innovation and Security**: While AI can drive innovation in infrastructure management, it is crucial to balance technological advancements with security concerns. This includes safeguarding data integrity, maintaining redundancy, and implementing fail-safe mechanisms to ensure system robustness against both natural and human-induced disruptions.

6. **Future Directions**: The intersection of AI and critical infrastructure resilience is a dynamic field, requiring continuous research and innovation. Prioritizing interdisciplinary approaches and fostering partnerships across sectors will be essential for addressing emerging challenges and maximizing the potential of AI in building resilient systems.

In conclusion, while AI offers many opportunities to enhance the resilience of critical infrastructure, its integration must be pursued with caution and foresight. By addressing ethical, societal, and risk management dimensions, stakeholders can use AI's potential to create infrastructure systems that are not only more robust and adaptive but also aligned with societal values and priorities.

APPENDIX: The use of AI application in Critical Infrastructure for Resilience, related risk and impacts on ethics and environmental issues *(Raffaella D'Alessandro)*

The integration of Artificial Intelligence in Critical Infrastructure is crucial for improving resilience, preventing disruptions, and ensuring the continuous and efficient functioning of these vital systems. Anyway the use of Artificial Intelligence in critical infrastructure can generate risk with impact on ethics and environmental issues.

The following table lists examples of use of AI application in main Critical Infrastructure in order to enhance resilience and depict related risk and impacts on ethics and environmental issues.

The Table was developed using ChatGPT v3 as a support for research activities.

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
Energy			
Electrical Grids	Predictive Maintenance	AI algorithms analyze data from sensors installed on the grid to predict equipment failures before they happen, reducing downtime. For example, identifying weaknesses in transformers or power lines to replace them before they cause outages.	Risks include data privacy concerns with extensive monitoring, potential job displacement, and the environmental impact of producing and disposing of sensors and AI hardware.
	Load Forecasting	Machine learning models predict energy demand, allowing for better load management and preventing blackouts. For example, adjusting energy production based on real-time consumption patterns.	Over-reliance on AI could lead to system vulnerabilities if the algorithms fail. There's also the risk of biased data leading to inaccurate predictions.
	Fault Detection and Isolation	AI systems quickly identify faults in the grid and isolate affected areas, minimizing service disruption. For example, detecting short circuits or overloads and rerouting power.	Ethical concerns include the transparency of AI decisions and potential biases in fault detection. The environmental impact involves the energy consumption of AI systems.
Oil and Gas Pipelines	Leak Detection	AI models analyze sensor data to detect leaks or irregularities in pipelines, allowing rapid response	Environmental risks include potential leaks if AI fails to detect them, and ethical

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
		to prevent environmental damage. For example, identifying small leaks that could lead to major spills.	concerns around data privacy.
	Predictive Maintenance	Monitoring the condition of pipelines to predict and schedule maintenance, avoiding failures. For example, analyzing pressure and flow data to schedule repairs.	Risks involve data security and the environmental impact of disposing old pipeline materials.
	Supply Chain Optimization	AI optimizes logistics and distribution, ensuring a steady supply. For example, predicting demand spikes and adjusting supply routes.	Ethical concerns include the potential loss of jobs and the environmental impact of increased transport efficiency leading to more frequent operations.
Energy Production Plants	Operational Efficiency	AI optimizes the performance of power plants by adjusting variables to maximize output and efficiency. For example, fine-tuning fuel combustion processes.	Risks include potential job displacement and the environmental impact of increased fossil fuel efficiency leading to higher emissions.
	Predictive Analytics	Predicting equipment failures and maintenance needs to avoid unplanned outages. For example, identifying signs of turbine weaknesses.	Ethical concerns include data privacy and security. Environmental risks involve the disposal of old equipment.
	Renewable Energy Integration	AI manages the variability of renewable energy sources by predicting generation and adjusting grid operations accordingly. For example, balancing solar and wind inputs with demand.	Risks include over-reliance on AI, which could fail, and the environmental impact of AI hardware. Ethical concerns involve data privacy.
Water			
Water Supply Systems	Leak Detection	AI systems monitor water flow and pressure to detect leaks in real-time, reducing water loss. For	Environmental impact involves the disposal of outdated water infrastructure. Ethical

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
collection, treatment and distribution in the territory		example, detecting leaks in underground pipes.	concerns include data privacy and security.
	Demand Forecasting	Predicting water demand to ensure adequate supply and efficient resource allocation. For example, adjusting water distribution based on weather forecasts and usage patterns.	Risks include inaccuracies leading to water shortages. Ethical concerns involve data privacy and potential biases.
	Quality Monitoring	AI analyzes water quality data to detect contamination early and ensure a safe water supply. For example, detecting pollutants or chemical imbalances.	Environmental impact involves the disposal of contaminated water. Ethical concerns include the accuracy and transparency of AI decisions.
	Desalination of marine waters for supply to the islands	AI analyzes water quality data to detect the right filtering for desalinization adopted and ensure a safe water supply.	Societal risk include potential harms to people, due to not safe water supply, if AI systems fail.
Wastewater Treatment Systems Wastewater disposal	Process Optimization	AI optimizes treatment processes to improve efficiency and reduce energy consumption. For example, adjusting chemical dosages in real-time.	Environmental risks include potential pollution if AI systems fail. Ethical concerns involve the transparency of AI decisions.
	Predictive Maintenance	Monitoring equipment to predict failures and schedule timely maintenance. For example, detecting weaknesses in pumps and filters.	Risks include data security and the environmental impact of disposing of old equipment. Ethical concerns involve potential job displacement.
	Anomaly Detection	Identifying unusual patterns in wastewater composition to detect issues such as industrial spills. For example, detecting high levels of toxic substances.	Environmental impact involves potential pollution if AI fails to detect anomalies. Ethical concerns include accuracy of AI decisions.

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
Transport			
Road and Highway Networks	Traffic Management	AI systems analyze traffic data to optimize traffic flow, reduce congestion, and improve safety. For example, adjusting traffic light timings based on real-time traffic conditions.	Ethical concerns include data privacy and the potential for biased decision-making. Environmental impact involves the increased energy consumption of AI systems.
	Predictive Maintenance	Monitoring road conditions to predict and schedule maintenance before serious deterioration occurs. For example, detecting pothole formation.	Risks involve data security and the environmental impact of construction waste. Ethical concerns include potential job displacement.
	Incident Detection	Using AI to detect accidents or road obstructions in real-time and manage emergency response. For example, detecting car crashes and alerting emergency services.	Ethical concerns include data privacy and the transparency of AI decisions. Environmental risks involve the response to hazardous material spills.
Railways	Operational Efficiency	AI optimizes train schedules and routing to improve efficiency and reduce delays. For example, adjusting schedules based on passenger demand.	Risks include over-reliance on AI, which could fail. Ethical concerns involve job displacement and data privacy. Environmental impact includes energy consumption of AI systems.
	Predictive Maintenance	Monitoring rail tracks and trains to predict maintenance needs and avoid breakdowns. For example, detecting weaknesses on rail tracks.	Risks involve data security and the disposal of old rail materials. Ethical concerns include job displacement.
	Safety Monitoring	AI systems monitor for safety hazards, such as obstructions on the tracks or mechanical failures. For example, detecting obstacles on the tracks.	Ethical concerns include the accuracy of AI decisions and data privacy. Environmental risks involve the response to hazardous materials.

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
Ports and Maritime Transport	Logistics Optimization	AI improves the efficiency of cargo handling and port operations, reducing delays. For example, optimizing container placement and movement.	Risks include job displacement and data privacy concerns. Environmental impact involves increased port activity and pollution.
	Vessel Traffic Management	AI helps manage ship traffic to prevent collisions and ensure smooth operations. For example, optimizing ship docking schedules.	Ethical concerns include the accuracy and transparency of AI decisions. Environmental risks involve potential marine pollution.
	Predictive Maintenance	Monitoring equipment and infrastructure to predict failures and schedule maintenance. For example, detecting weaknesses on cranes and docking equipment.	Risks involve data security and the disposal of old equipment. Ethical concerns include job displacement.
Airports and Civil Aviation	Air Traffic Management	AI optimizes flight paths and schedules to reduce delays and improve safety. For example, optimizing air traffic control operations.	Risks include over-reliance on AI, which could fail. Ethical concerns involve data privacy and the transparency of AI decisions. Environmental impact includes increased air traffic and emissions.
	Passenger Flow Management	AI analyzes passenger data to optimize security and boarding processes, reducing wait times. For example, adjusting security checkpoint staffing based on real-time data.	Ethical concerns include data privacy and potential biases in AI decisions. Environmental risks involve increased energy consumption of AI systems.
	Predictive Maintenance	Monitoring aircraft and airport infrastructure to predict and prevent failures. For example, detecting weaknesses in aircraft engines.	Risks involve data security and the disposal of old aircraft parts. Ethical concerns include job displacement.
Communications			

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
Telecommunications Networks	Network Optimization	AI optimizes network performance, managing bandwidth and ensuring consistent service quality. For example, dynamically adjusting network traffic routing.	Ethical concerns include data privacy and the potential for biased decision-making. Environmental impact includes the increased energy consumption of AI systems.
	Anomaly Detection	Identifying unusual patterns that may indicate cyber attacks or technical issues. For example, detecting unusual data traffic patterns.	Ethical concerns include data privacy and security. Environmental risks involve the energy consumption of AI systems.
	Predictive Maintenance	Monitoring network components to predict and address potential failures. For example, detecting weaknesses in network hardware.	Risks involve data security and the disposal of old network equipment. Ethical concerns include job displacement.
Internet Infrastructure	Load Balancing	AI ensures efficient distribution of internet traffic to prevent overloads and downtime. For example, adjusting server loads based on real-time data.	Ethical concerns include data privacy and the transparency of AI decisions. Environmental impact involves the energy consumption of AI systems.
	Cybersecurity	AI systems detect and respond to cyber threats in real-time, protecting the integrity of internet infrastructure. For example, identifying and mitigating DDoS attacks.	Ethical concerns include the accuracy and transparency of AI decisions. Environmental risks involve the energy consumption of AI systems.
	Service Quality Management	Monitoring and optimizing service delivery to ensure high-quality user experiences. For example, adjusting streaming quality based on network conditions.	Ethical concerns include data privacy and potential biases in AI decisions. Environmental risks involve the energy consumption of AI systems.
Radio and Television Transmission Systems	Signal Optimization	AI ensures optimal signal quality and coverage, reducing interruptions. For example,	Ethical concerns include data privacy and the transparency of AI decisions. Environmental

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
		dynamically adjusting transmission power.	impact includes the energy consumption of AI systems.
	Predictive Maintenance	Monitoring transmission equipment to predict and schedule maintenance. For example, detecting weaknesses in transmission towers.	Risks involve data security and the disposal of old transmission equipment. Ethical concerns include job displacement.
	Content Delivery Optimization	AI optimizes the delivery of content to ensure consistent quality and reduce buffering. For example, adjusting content delivery networks based on user demand.	Ethical concerns include data privacy and potential biases in AI decisions. Environmental risks involve the energy consumption of AI systems.
Finance			
Banking Systems	Fraud Detection	AI analyzes transaction data to detect and prevent fraudulent activities. For example, identifying unusual spending patterns.	Ethical concerns include data privacy and the accuracy of AI decisions. Environmental impact includes the energy consumption of AI systems.
	Risk Management	AI models assess financial risks and optimize investment strategies. For example, predicting market trends.	Ethical concerns include the potential for biased decision-making and data privacy. Environmental risks involve the energy consumption of AI systems.
	Customer Service	AI-powered chatbots and virtual assistants improve customer service by handling inquiries and transactions. For example, answering customer questions in real-time.	Ethical concerns include data privacy and potential job displacement. Environmental impact includes the energy consumption of AI systems.
Stock Exchanges	Algorithmic Trading	AI algorithms execute trades based on market data, optimizing returns and reducing risks. For example, making split-second trading decisions.	Ethical concerns include market manipulation and the transparency of AI decisions. Environmental

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
			impact involves the energy consumption of AI systems.
	Market Analysis	AI analyzes market trends and data to provide insights for better decision-making. For example, predicting stock price movements.	Ethical concerns include data privacy and the accuracy of AI predictions. Environmental risks involve the energy consumption of AI systems.
	Security Monitoring	AI systems detect and respond to security threats, protecting the integrity of trading systems. For example, identifying unusual login patterns.	Ethical concerns include data privacy and the transparency of AI decisions. Environmental impact includes the energy consumption of AI systems.
Health			
Hospitals and Clinics	Predictive Analytics	AI predicts patient admission rates, optimizing resource allocation. For example, forecasting seasonal flu outbreaks.	Ethical concerns include data privacy and potential biases in AI predictions. Environmental impact involves the energy consumption of AI systems.
	Diagnostic Assistance	AI assists doctors in diagnosing diseases by analyzing medical data. For example, identifying early signs of cancer in medical images.	Ethical concerns include the accuracy and transparency of AI diagnoses. Environmental risks involve the energy consumption of AI systems.
	Operational Efficiency	AI optimizes hospital operations, from scheduling to inventory management. For example, optimizing staff schedules based on patient loads.	Ethical concerns include data privacy and potential job displacement. Environmental impact includes the energy consumption of AI systems.
Emergency Health Systems	Resource Allocation	AI ensures optimal allocation of emergency resources, such as ambulances and medical personnel. For example, deploying	Ethical concerns include data privacy and the transparency of AI decisions. Environmental

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
		ambulances based on real-time traffic data.	risks involve the energy consumption of AI systems.
	Real-Time Monitoring	AI monitors health data to detect outbreaks and respond to health emergencies. For example, identifying spikes in flu cases.	Ethical concerns include data privacy and potential biases in AI decisions. Environmental impact involves the energy consumption of AI systems.
	Telemedicine	AI enhances telemedicine services, providing remote diagnostics and consultations. For example, analyzing patient data to provide medical advice.	Ethical concerns include data privacy and the accuracy of AI diagnoses. Environmental risks involve the energy consumption of AI systems.
Food Sector			
Food Production and Distribution	Supply Chain Optimization	AI optimizes the food supply chain, from production to distribution, reducing waste and ensuring food security. For example, predicting harvest yields and adjusting distribution accordingly.	Ethical concerns include data privacy and potential job displacement. Environmental impact involves the energy consumption of AI systems.
	Quality Control	AI monitors food quality and safety, detecting contamination or spoilage. For example, identifying bacterial contamination in food products.	Ethical concerns include the accuracy and transparency of AI decisions. Environmental risks involve the disposal of contaminated food.
	Demand Forecasting	AI predicts food demand, helping producers and retailers manage inventory and reduce waste. For example, forecasting consumer demand for seasonal products.	Risks include inaccuracies leading to food shortages or surpluses. Ethical concerns involve data privacy and potential biases. Environmental impact includes the energy consumption of AI systems.
Security			

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
Law Enforcement	Crime Prediction	AI analyzes data to predict and prevent criminal activities. For example, identifying crime hotspots.	Ethical concerns include potential biases and the transparency of AI decisions. Environmental impact involves the energy consumption of AI systems.
	Surveillance	AI enhances surveillance systems with facial recognition and behavioral analysis. For example, identifying suspects in public areas.	Ethical concerns include privacy violations and potential biases in AI systems. Environmental impact includes the energy consumption of AI systems.
	Resource Allocation	AI optimizes the deployment of law enforcement resources. For example, adjusting patrol routes based on crime data.	Ethical concerns include data privacy and the transparency of AI decisions. Environmental risks involve the energy consumption of AI systems.
National Defense	Threat Detection	AI systems detect and analyze potential threats to national security. For example, identifying cyber attacks.	Ethical concerns include the accuracy and transparency of AI decisions. Environmental impact involves the energy consumption of AI systems.
	Cybersecurity	AI protects defense infrastructure from cyber attacks. For example, detecting and mitigating hacking attempts.	Ethical concerns include data privacy and the accuracy of AI systems. Environmental impact includes the energy consumption of AI systems.
	Operational Planning	AI supports strategic and tactical planning for military operations. For example, simulating battlefield scenarios.	Ethical concerns include the transparency of AI decisions and potential biases. Environmental impact involves the energy consumption of AI systems.
Civil Protection	Disaster Response	AI improves emergency response by predicting natural disasters and optimizing evacuation plans. For	Ethical concerns include data privacy and the accuracy of AI predictions.

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
		example, forecasting hurricanes and planning evacuations.	Environmental impact involves the energy consumption of AI systems.
	Resource Management	AI ensures optimal allocation of resources during emergencies. For example, distributing relief supplies based on real-time needs.	Ethical concerns include data privacy and the transparency of AI decisions. Environmental risks involve the energy consumption of AI systems.
	Risk Assessment	AI assesses risks and improves disaster preparedness. For example, identifying areas prone to flooding.	Ethical concerns include data privacy and potential biases in AI decisions. Environmental impact involves the energy consumption of AI systems.
Public Services			
Waste Management Systems	Route Optimization	AI optimizes waste collection routes, reducing operational costs and environmental impact. For example, adjusting routes based on real-time waste levels.	Ethical concerns include data privacy and potential job displacement. Environmental impact involves the energy consumption of AI systems.
	Recycling Management	AI improves recycling processes and efficiency. For example, sorting recyclable materials automatically.	Ethical concerns include data privacy and potential biases in AI decisions. Environmental risks involve the disposal of non-recyclable waste.
	Waste Monitoring	AI monitors waste levels and predicts collection needs. For example, detecting overflowing bins.	Ethical concerns include data privacy and the accuracy of AI predictions. Environmental impact includes the energy consumption of AI systems.
Civil Infrastructure	Building Management	AI optimizes the management and maintenance of public buildings. For example, adjusting heating	Ethical concerns include data privacy and potential job displacement. Environmental impact

Critical Infrastructure	AI Application for Resilience	Description	Ethical and Environmental Risks/Impacts
		and cooling systems based on occupancy.	involves the energy consumption of AI systems.
	Energy Efficiency	AI improves the energy efficiency of public infrastructure. For example, optimizing lighting based on natural light levels.	Ethical concerns include data privacy and the transparency of AI decisions. Environmental risks involve the energy consumption of AI systems.
	Safety Monitoring	AI systems monitor and ensure the safety of public buildings. For example, detecting structural weaknessesnesses.	Ethical concerns include the accuracy and transparency of AI decisions. Environmental impact involves the energy consumption of AI systems.

AUTHORS

(*in alphabetical order*)

Silvano Bari

Graduated in Statistical and Demographic Sciences and Master's Degree in Computer Law, former Head of Security and Privacy of Alitalia - Italian Airlines. Today he is a professor of "Risk Assessment" at the "Campus Bio-medico di Roma" University and vice president of AIIC (Italian Association of Critical Infrastructure Experts). It is CISM certified.

Glauco Bertocchi

Degree in Physics at the University of Rome "la Sapienza" More than 40 years of experience in IT and security acquired within universities and national institutions. Active in research in the field of protection and resilience of critical infrastructure. He coordinates a development and research group of ISACA Rome for the application of quantitative methods in the analysis of cyber risks and more. Vice-President of the ISACA Rome chapter, member of the AIIC Board of Directors.

Sandro Bologna

Graduated in Physics at the Sapienza University of Rome. The main research activity concerns Critical Infrastructure Resilient to Disasters of different nature, with particular attention to the modeling, simulation and analysis of vulnerabilities and interdependencies, as well as the use of different technologies in order to increase their security. Recently he has started working on the topic of Sustainable Development Indices with reference to the Sustainable Development Goals of the 2030 Agenda.

Luigi Carrozzi

(CGEIT, CRISC, Auditor L.A. ISMS, ISMS Senior Manager, Privacy Officer) Graduated in Statistical Sciences, he has over 35 years of experience in ICT Governance, Risk Management and Compliance at leading private and public organizations.

Alberto Caruso de Carolis

A retired senior officer of the Guardia di Finanza, since 2002 he has been a company manager, in the airport sector, in senior management, *security*, *internal audit* and public/private relations and synergies management, also in the Cybersecurity sector; he is currently a partner in strategic consulting companies and lecturer at the First Level Master " *Crisis & Disaster Management"* at the Università Cattolica del Sacro Cuore in Milan.

Raffaella D'Alessandro

Degree in Economics. 40 years of experience in ICT, of which 35 years in Cybersecurity, Data Protection, Management Systems, Standards and Digital Law Compliance. She worked at Olivetti, Arthur Andersen, Ernst & Young and IBM. Member of the Board of Directors of the Italian Association of Critical Infrastructure Experts, Founding Member and Secretary of TOPForGrowth, Official Speaker Word Protection Forum, past member of UNI - UNINFO.

Francesca Della Mea

Graduated in Business Economics and Organization from Bocconi University in Milan, with more than 20 years in the Consulting field with a specific focus on IT Security, and with extensive experience in managing complex transformation projects. Passionate about IT Risk Management, hold several cybersecurity certifications.

Luisa Franchina

Co-founder of AIIC, he is currently its President. She was Director General of the Secretariat for Critical Infrastructure (Presidency of the Council of Ministers 2010-2013). He has published numerous articles and books on the security and protection of critical infrastructure.

Adriana Peduto

Lawyer, partner of the E-lex firm, he deals with privacy and new technology law, with many years of experience in copyright and consumer law. ISO/IEC 27001:2022 Lead Auditor, she is Data Protection Officer (DPO) of public institutions and private companies. He deals with ethics & privacy compliance profiles in EU research projects, in the fields of cyber security of critical infrastructure and data energy management.

Giorgio Pizzi

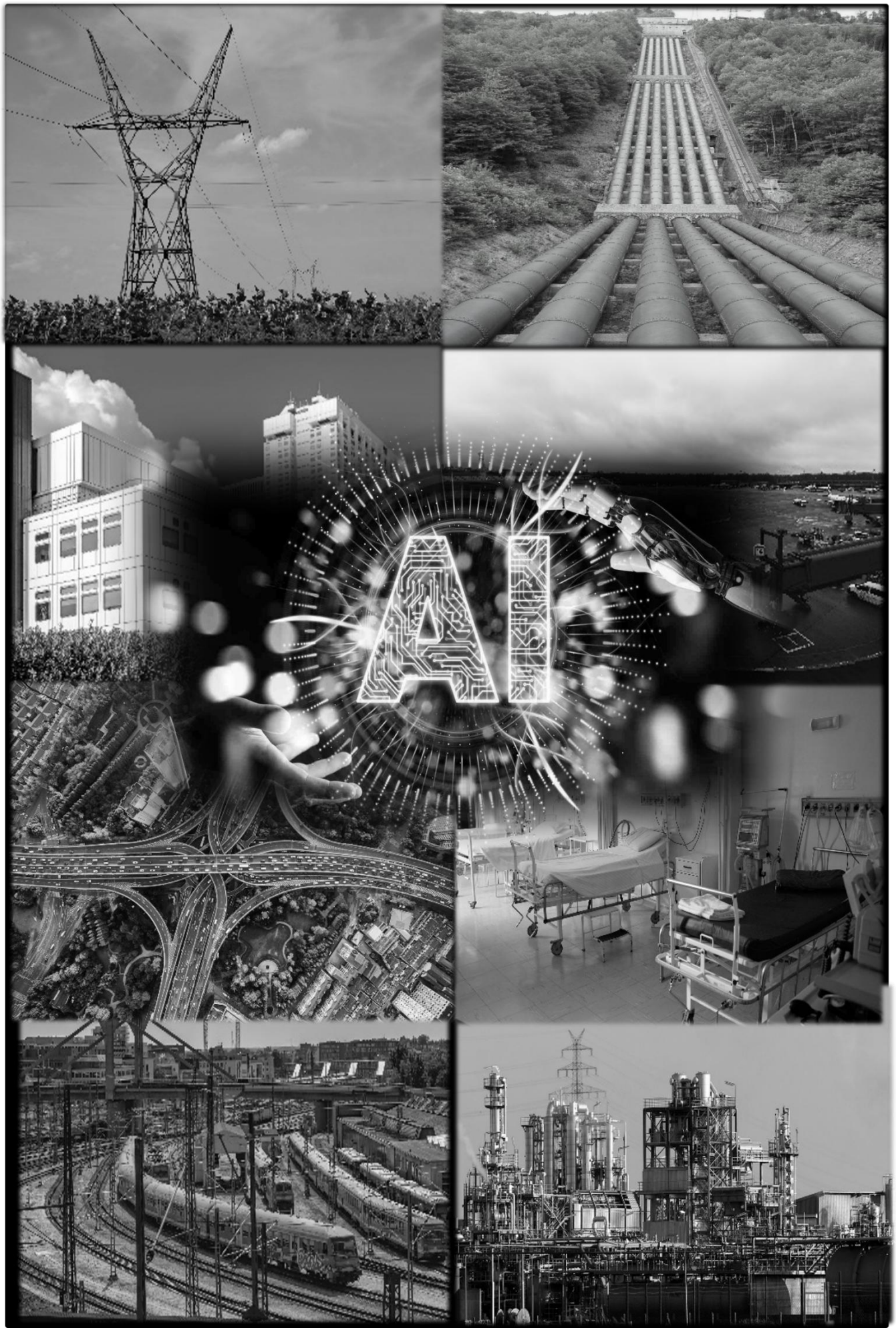
Graduated in Electronic Engineering, he is a manager of the Ministry of Infrastructure and Transport. He currently deals with digital platforms for mobility and is a member of various working groups and committees at ministerial, CEN and UITP level concerning the safety of transport systems, cableway technical standardization and the integration between cybersecurity and safety in transport systems.

Alberto Stefanini

Graduated in Electronic Engineering and Master in Classical Literature. His career spans research centers and electricity sector companies, including work with the Joint Research Center of the European Community on cybersecurity and nuclear waste classification. Since 2008 he has shaped and worked on several international collaborative projects under various funding schemes. His recent interests focus on AI applications and emerging technology economics.

Maria Beatrice Versaci

She holds a master's degree in Oriental Languages and Civilizations (Arabic) from La Sapienza University in Rome, then specialized in Strategic Protection of the Country System (Cyber Intelligence, Big Data and Security of Critical Infrastructures) at the Italian Society for International Organization (SIOI).


AI