ATLANTIS

In collaboration with SUNRISE

Securing Europe Critical Infrastructure:

Aligning AI innovation with policy and governance for Critical Infrastructure resilience

Policy Brief - August 2025

Co-funded by
the European Union

Executive Summary

As Artificial Intelligence (AI) becomes embedded in Europe's Critical Infrastructure (CI), its robust deployment has become a strategic imperative. Developed under the ATLANTIS project, this policy brief explores how emerging AI technologies, especially non-deterministic models, can be integrated into CI systems in alignment with the Network and Information Systems (NIS2), Critical Entities Resilience (CER) Directives, and AI Act frameworks to enhance resilience and reduce systemic risks. Across the European Union (EU), CI operators face mounting challenges from cyber threats, hybrid attacks, and geopolitical volatility. In response, strategic initiatives such as the Security Union Strategy, the NIS2 Directive, and the CER Directive reinforce the political priority of safeguarding essential services.

Within this evolving landscape, AI is both a force multiplier and a potential risk vector. While AI applications are already improving predictive maintenance, incident detection and operational performance, the deployment of opaque models such as deep learning systems or large language models raises new concerns around trust, traceability, and cascading vulnerabilities.

This brief recommends concrete steps to ensure secure AI adoption in CI, including mandatory system disclosure and Software Bill of Materials (SBOM) practices (see definition and role in the Recommendations section), to be operationalized through delegated legislation or sectoral guidance under the AI Act, NIS2, and Cyber Resilience Act (CRA). It also calls for cross-sector guidance and the creation of a dedicated EU AI Resilience Lab to test and validate high-risk AI systems in CI environments.

Three core challenges are identified:

- Limited transparency regarding the origins and components of pre-trained AI models, including their data and software supply chains.

- Regulatory uncertainty due to overlapping obligations under the AI Act, NIS2 and the CRA, making compliance complex for CI operators.

- Increased vulnerability to AI-driven attacks and cascading failure scenarios across interconnected systems.

Pre-trained AI models pose challenges in CI due to limited transparency over initial training data and methodologies. As highlighted by ATLANTIS and SUNRISE, understanding how AI systems are trained is essential for assessing reliability and trustworthiness in high-stakes environments. The ATLANTIS Guidelines on Pre-Trained Models further elaborate on these issues, providing recommendations to improve transparency and risk awareness for CI operators.

Key insights include:

- **Not all AI systems are equally robust:** Complex, non-explainable models pose higher risks and, if used in CI, must be subject to enhanced oversight and safeguards.
- **Trust in AI must be systemic, requiring assurance across the full chain:** This ranges from model quality and training data to vendor practices, software components (SBOM) and governance frameworks.
- **Regulation is strategically but not yet operationally converging:** NIS2, CER and the AI Act offer a shared policy foundation but require operational interoperability and guidance.

The brief outlines short-, medium-, and long-term recommendations to:

- Mandate AI system disclosure and SBOM use across CI sectors.
- Issue joint EU guidance on the deployment of high-risk AI in essential services.
- Launch an EU AI Resilience Lab and explore certification schemes for trustworthy AI-driven CI tools.

This document is intended to inform a broad range of stakeholders involved in the integration of AI into CI, including policymakers, national authorities, CI operators, regulators, and technology developers. The aim is to support actionable implementation, regulatory alignment, and the development of trustworthy AI practices tailored to the operational realities of Europe's critical systems.

Disclaimer:

The views and recommendations expressed in this policy brief are those of ATLANTIS project and contributors and do not necessarily reflect the official position, mandate, or institutional agendas of any participating organization.

The work presented in this policy brief has been partially funded by the ATLANTIS project, which has received funding from the European Union's Horizon Europe framework programme under grant agreement No. 101073909.

The work presented in this document represent the views of the authors only. The European Research Executive Agency and the European Commission are not responsible for any use of the included information.

Introduction

Europe’s CI is increasingly exposed to complex and evolving hybrid threats. In this context, AI has emerged as a powerful enabler for resilience, offering capabilities such as predictive maintenance, real-time system monitoring, and improved crisis response. To clarify what constitutes an AI system, this brief adopts the definition provided in the AI Act under Article 3(1):

“AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”.

Also at the WP level, the operational adoption of innovative technology for CI resilience validated assumptions about normal or emergency operations, with a limited number of in-between situations or scenarios where collaboration and adaptivity depend on workforce and supply chain availability.

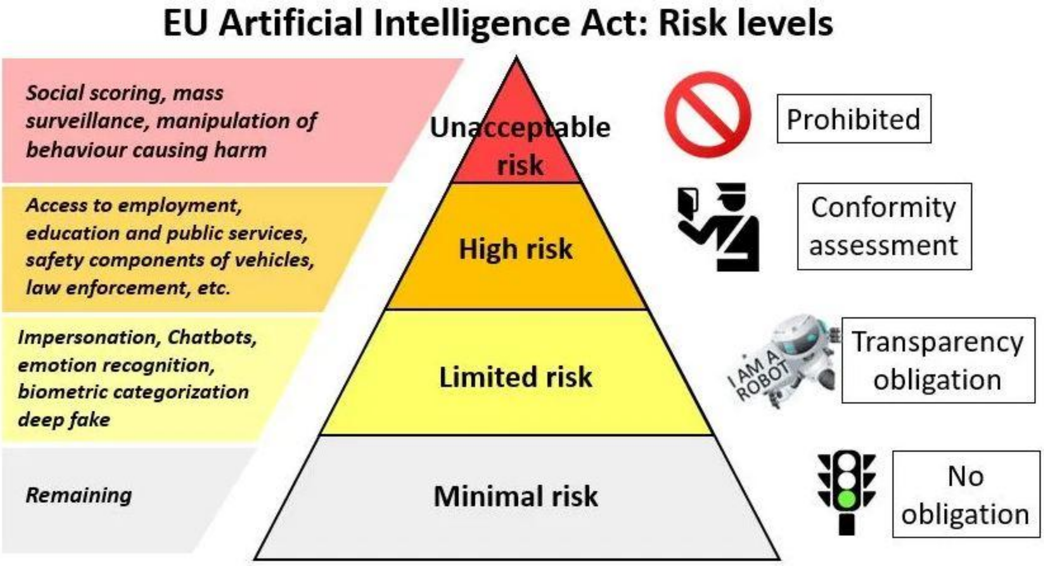

EU Artificial Intelligence Act: Risk levels
Social scoring, mass surveillance, manipulation of behaviour causing harm
Access to employment, education and public services, safety components of vehicles, law enforcement, etc.
Impersonation, Chatbots, emotion recognition, biometric categorization deep fake
Remaining
Unacceptable risk
High risk
Limited risk
Minimal risk
Prohibited
Conformity assessment
I AM A ROBOT
Transparency obligation
No obligation

Figure 1: *The four risk levels under AI Act*

As illustrated in Figure 1, the AI Act distinguishes between four risk levels, with high-risk systems requiring enhanced transparency, oversight, and risk mitigation measures.

This brief focuses on these high-risk, non-deterministic AI applications within CI environments, especially those used for remote monitoring, decision support, and infrastructure management.

Yet, alongside these benefits, the deployment of AI in CI environments raises important concerns around safety, transparency, and regulatory compliance. The ATLANTIS project addresses these challenges by testing how emerging AI systems can align with EU Directives e.g. the CER and NIS2 Directives, as well as the AI Act Regulation.

High-risk AI systems deployed in CI environments must meet strict requirements as defined by EU regulations, notably the AI Act. ATLANTIS pilot simulations and Policy Taskforce discussions have reinforced the operational relevance of these obligations and highlighted practical challenges in their implementation. Furthermore, ATLANTIS contributes to bridging the gap between AI regulation and operational implementation, offering a testing ground for AI-enabled risk detection, supply chain validation via SBOM, and operational mechanisms to build trust in CI contexts, e.g. explainability logs, audit trails and human-in-the-loop safeguards.

Complementary initiatives like the SUNRISE project further explore how trust mechanisms (e.g. transparent AI models for decision support, credibility scoring of information sources, human-in-the-loop crisis dashboards) can shift risk perceptions in crisis settings, such as pandemics. These tools developed by SUNRISE play a direct role in decision-making and must therefore meet rigorous trustworthiness requirements, particularly when their output influences critical decisions.

The definition adopted from the AI Act (Article 3(1)) refers to AI systems that may operate with varying levels of autonomy and adaptiveness. In this context, the brief focuses on AI systems that generate outputs which may vary over time for the same input, i.e. they are non-deterministic in nature. These adaptive AI systems evolve based on new data or changing conditions, making their output unpredictable and difficult to reproduce consistently.

In contrast, deterministic and explainable models, which are built on known data, have clear logic, and can be logged and audited. They pose significantly lower risk. They reflect standard, observable behaviors and can be verified using traditional testing and validation methods. These are still considered AI under the AI Act, but they present limited challenges from a compliance and trust standpoint.

However, non-deterministic or unexplainable models introduce higher risk. Their output may be unpredictable, making it difficult to fully assess their performance or consequences. Under the AI Act, such models, i.e. when used in high-risk contexts like CI, must meet stricter requirements, as outlined in Articles 9 to 15, including:

- The use of high-quality, unbiased training data (Article 10).
- Comprehensive technical documentation (Article 11).
- Robust testing, validation, and monitoring procedures (Articles 9, 12, and 15).

Yet, even with rigorous testing, complete certainty about performance may not be possible. Therefore, it is crucial that CIs are informed about the nature of the AI models embedded in their tools, whether they are deterministic or adaptive.

This requirement extends further to the entire software supply chain. Ensuring trustworthiness requires scrutiny not only of the AI models themselves and the underlying software components they rely on, but also of the training data and datasets used throughout the AI lifecycle, including their provenance, representativeness, and potential biases. Within ATLANTIS, particular attention has been paid to the traceability of these components through SBOMs, a key enabler of supply chain trust. The chain of trust must be traceable and reliable, since the inclusion of stochastic or opaque components can compromise overall system integrity. A stable and verifiable SBOM is therefore essential.

The EU regulatory response to these needs includes:

- **The Cyber Resilience Act (CRA):** Governing the software supply chain and SBOM practices.
- **The AI Act:** Applying a risk-based approach to the use of AI components in tools.

These regulations aim to ensure that AI-powered systems in CI are robust, secure, transparent, and comply with EU standards. The following focuses on unexplainable, non-deterministic AI, including models such as Large Language Models, Deep Learning models, and other similarly complex systems.

AI for critical infrastructure Resilience: Use cases and strategic opportunities

AI has the potential to support systems that continuously adapt by enabling real-time information exchange across layers of governance, operations, and monitoring. For this potential to be fully realized in the context of CI, governance approaches must evolve in parallel. The goal is to enhance the precision of risk anticipation and response while avoiding burdensome complexity. This principle, already outlined in ATLANTIS policy work and echoed in SUNRISE crisis coordination activities, emphasizes the value of operationally grounded and forward-looking governance frameworks.

Findings from both ATLANTIS and SUNRISE underline the importance of structured information sharing, across sectors (horizontally) and between governance levels (vertically). This includes integrating traditional data sources, AI-generated outputs, and even real-time human observations. However, aligning these inputs remains a challenge, especially where public and private entities operate with different mandates. This was evident during the COVID-19 pandemic and has been addressed through multi-stakeholder exercises in both projects, highlighting the need to plan such coordination in advance.

Crisis management should be understood as a socio-technical system that relies on feedback loops, adaptive decision-making, and often, permission-based access to information. Exercises conducted under ATLANTIS and SUNRISE projects have shown how pre-configured response protocols supported by AI tools can significantly enhance agility in complex, high-pressure situations.

Trust in AI is crucial but must extend beyond algorithms. A major challenge stems from pre-trained models, which often obscure their original training data and methodologies. This opacity limits CI operators' ability to assess model reliability or detect bias. Both ATLANTIS and SUNRISE underscored that understanding how a system was trained is operationally critical when AI is deployed in high-stakes environments such as infrastructure protection.

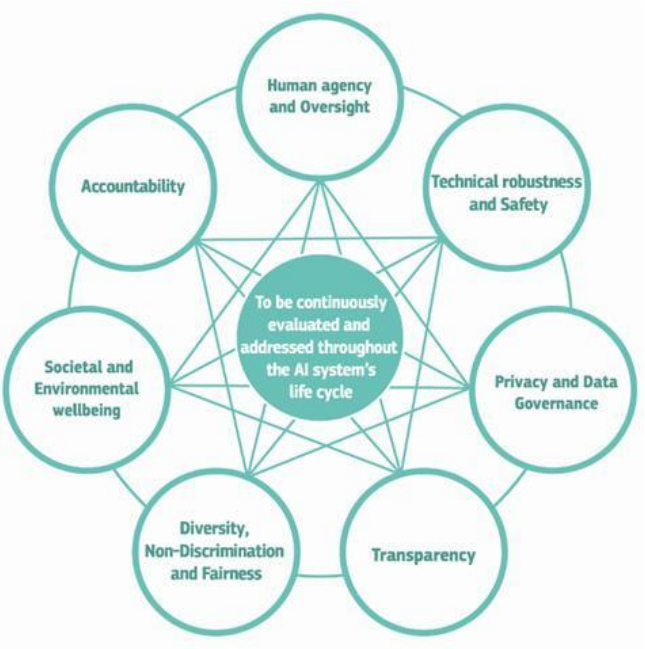

Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

Figure 2: *The seven trustworthiness requirements defined by the European Commission's Ethics Guidelines for Trustworthy AI (evaluated continuously through the system's lifecycle).*

Remote monitoring tools, for example, increasingly rely on satellite imagery or drone footage processed by AI. To ensure accountability, such tools must allow for periodic review of archived data and provide transparency about how decisions were reached. In ATLANTIS use cases, emphasis was placed on documenting potential AI limitations and helping end-users anticipate system errors or detection gaps.

Given the rapid development and deployment of AI technologies across CI sectors, the EU has responded with a comprehensive legislative framework under the AI Act. This regulation introduces a risk-based classification system, ranging from **unacceptable** to **high, limited**, and **minimal** risk (see Figure 1), and places clear obligations on providers of high-risk AI systems, such as those used in the management and operation of CI. These obligations include maintaining detailed technical documentation of training processes, evaluation results, and intended use-cases, to be made available upon request to the AI Office and competent national authorities.

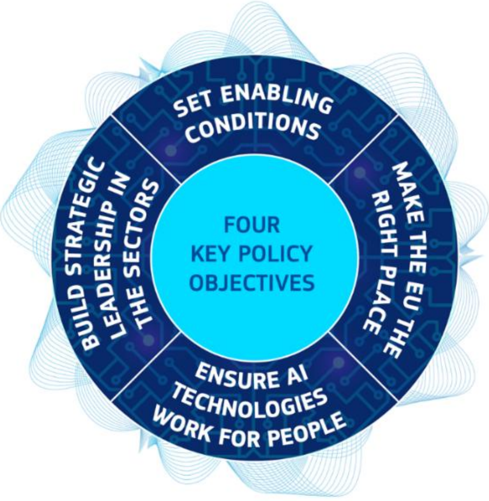

SET ENABLING CONDITIONS
MAKE THE EU THE RIGHT PLACE
ENSURE AI TECHNOLOGIES WORK FOR PEOPLE
BUILD STRATEGIC LEADERSHIP IN THE SECTORS
FOUR KEY POLICY OBJECTIVES

Figure 3: *The four key policy objectives guiding the European Commission's AI Strategy, each supporting the deployment of AI technologies in a trustworthy, human-centric and resilient manner across sectors, including CI.*

At the same time, the EU continues to invest in future-oriented AI ecosystems that support industrial competitiveness, digital sovereignty, and operational resilience. AI is already driving innovation across European sectors, ranging from healthcare and communications to climate monitoring and space applications, i.e. by enabling improved forecasting, early warning, and infrastructure management capacities. In this context, both ATLANTIS and SUNRISE have illustrated how AI can reinforce coordination and crisis response through predictive analytics, sensor integration, and real-time decision support tools.

To bolster Europe's capacity in developing and scaling trustworthy AI, the **European High Performance Computing Joint Undertaking** (EuroHPC JU) is leading the establishment of a pan-European network of AI Factories, i.e. state-of-the-art facilities equipped with supercomputers, data centers, and support services. These sites will allow researchers and industry actors to test and refine large-scale models for high-risk applications, including CI. As of March 2025, 13 AI Factory sites have been selected across 11 Member States (MSs), including locations in Finland, Germany, Greece, France, and Slovenia.

Nevertheless, deploying AI within CI environments requires scrutiny. A key challenge identified across both ATLANTIS and SUNRISE use cases is the limited visibility of the internal workings of pre-trained models, particularly when the original training data is unknown. This lack of transparency can introduce operational uncertainty and undermine trust in automated decision-making tools. Moreover, as highlighted in ATLANTIS threat modelling work, remote inspection systems using pre-trained models, whether for processing satellite imagery or drone footage, must be accompanied by traceable inputs, audit logs, and mechanisms for post-deployment validation. In SUNRISE, both open vocabulary models and Vision-Language Models (VLMs) are used as-is with open weights. Their training was conducted by the third-party institutions using large-scale public datasets (e.g., LAION-5B, OpenImages) but they are not retrained in the project.

Security concerns extend beyond technical transparency. Even models used solely for classification without retraining can be vulnerable to supply chain threats. Malicious actors may exploit early-stage training processes to embed subtle biases or behavioral anomalies, i.e. a risk known as data poisoning. Such vulnerabilities can persist even after fine-tuning and may manifest during live operations. Ensuring that CI operators have full visibility of the provenance, limitations, and evaluation results of AI systems is therefore essential.

To mitigate these risks and foster informed deployment, ATLANTIS recommends documenting the full lifecycle of AI tools integrated into CI workflows. While Article 11 of the AI Act mandates technical documentation for high-risk AI systems, ATLANTIS recommends extending this approach to cover the full operational lifecycle of AI tools used in CI. This includes dynamic updates, monitoring procedures, and documentation tailored for use by CI operators and crisis managers, i.e. ensuring that compliance obligations also translate into practical resilience and risk awareness. End-users must be aware of potential errors and system limitations and be equipped with procedures for monitoring and review.

As AI becomes more deeply embedded into Europe's CI ecosystem, ongoing collaboration between policymakers, researchers and CI stakeholders, as fostered through ATLANTIS and SUNRISE, will be crucial for aligning innovation with resilience.

Risks and limitations of AI for critical infrastructure Resilience

As AI and machine learning (ML) systems become increasingly embedded in the protection and operation of CI, their associated risks must be rigorously assessed. These risks stem from (1) **internal integration** by CI operators, sometimes without full visibility into system behavior, and (2) **supply chain vulnerabilities**, where AI components from third parties may lack transparency and reliability. While adversarial AI is an emerging concern (e.g. AI-generated cyberattacks or disinformation), this brief focuses on securing the safe and trustworthy deployment of AI within CI environments.

Internal operational risks relate to how AI/ML tools are integrated into CI workflows. Risks may arise not only from the technical performance of AI systems but also from how they are used by operators within dynamic or constrained environments. These constraints include limited computational resources, system dependencies, or temporary conditions, e.g. those observed during the COVID-19 pandemic. In SUNRISE, for instance, custom-trained models for tasks such as crack detection, corrosion, and fire were re-trained on high-performance GPUs, with training times ranging from several hours depending on the complexity of the dataset (which was sourced online from public repositories). However, for inference, i.e. real-time anomaly detection on UAVs operating at the edge, these models must be optimized to reduce hardware requirements, particularly when using Vision-Language Models (VLMs). These performance trade-offs between training and deployment can introduce risks, especially in resource-constrained environments. Importantly, even minor model errors can scale into severe consequences when propagated across complex decision chains. This highlights the need for deploying only validated and reliable AI systems, supported by robust risk assessment (RA) processes from the design stage through real-world operations.

Supply chain risks concern the integration of AI within broader CI system components. The AI lifecycle, i.e. from model development to deployment and maintenance, relies on input from multiple stakeholders. However, developers may not disclose the risk metrics or methodologies used, creating blind spots. This calls for a comprehensive supply chain RA approach, covering:

- AI/ML service and data provision contracts
- Security assessments of third-party vendors
- Verification of software and data authenticity
- Implementation SBOMs
- The use of hardware-enabled security and Chains of Trust (CoT)

AI governance in this domain should also account for dynamic updates and reconfigurations during crisis conditions. From a systemic perspective, security strategies must encompass the AI model itself, its features and inputs, the software infrastructure, and the hardware it runs on. RA should evolve across the AI deployment lifecycle, as both the nature and impact of risks change over time.

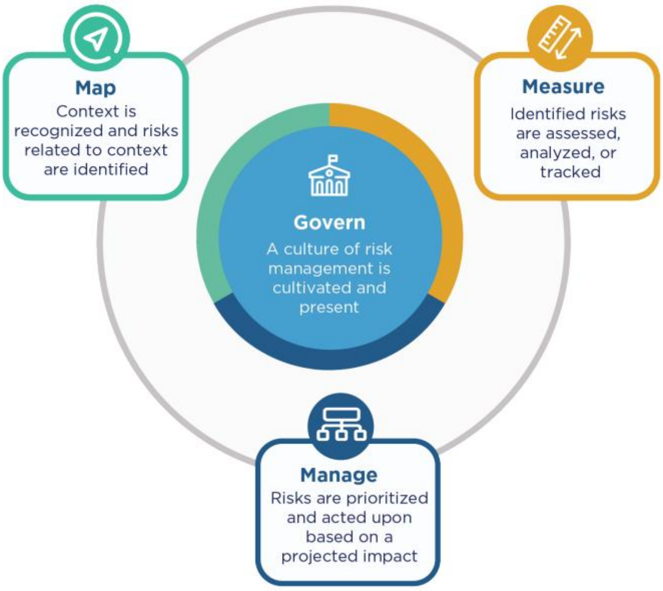

Map
Context is recognized and risks related to context are identified
Measure
Identified risks are assessed, analyzed, or tracked
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact

Figure 4: *Multi-source AI risk landscape across the CI lifecycle. Adapted from NIST AI RMF (2023).*

SUNRISE has underlined the importance of multi-stakeholder involvement in AI RA. Inputs from AI designers, domain experts, CI operators, and executive-level stakeholders help identify overlooked risks and align trust criteria across sectors. In some scenarios, especially under exceptional conditions, trade-offs between accuracy and privacy may need to be explicitly addressed.

Recent work in ATLANTIS has supported sector-wide interoperable RA frameworks to assess AI suitability in CI. This aligns with the National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework (AI RMF), which defines trustworthy AI as encompassing validity, reliability, security, safety, and resilience.

Based on its pilot simulations and policy taskforce discussions, ATLANTIS has identified three operational insights to inform AI governance for CI:

1. Non-explainable AI models pose unacceptable risks in high-stakes environments.
2. Trustworthiness must span the full supply chain, beyond isolated models.
3. CI operators require actionable guidance to apply EU regulatory frameworks.

Furthermore, **enhanced explainability** can be supported through **Chain of Thought** (CoT) approaches, allowing AI systems to present transparent step-by-step reasoning. This helps security analysts understand the AI's logic and improves decision-making during normal operations, incident response, and system audits.

Ultimately, managing AI risk in CI requires clear metrics, transparent governance, defined responsibilities, and collaboration across the ecosystem, i.e. from developers to end-users. The work of both projects illustrates how such a collaborative approach can translate to real-world impact.

Regulatory alignment and strategic implementation

The integration of AI into CI systems brings clear operational benefits, i.e. improving predictive maintenance, remote monitoring, anomaly detection, and data-informed decision-making. However, these capabilities also heighten exposure to risk, particularly in cybersecurity, operational integrity, and responsible automation. To address this dual challenge, the European regulatory landscape has progressively aligned around three cornerstone instruments: the **NIS2 and CER Directives**, and **the AI Act**. Each regulation focuses on a specific risk dimension, i.e. cybersecurity, physical/operational resilience, and AI governance. However, taken together: they create a solid, interlocking framework to support the secure and resilient deployment of AI across essential sectors.

The interplay between NIS2 and the AI Act is particularly relevant for CI operators. NIS2 sets robust cybersecurity obligations across sectors such as healthcare, digital infrastructure, and water management, requiring risk-based approaches and timely incident reporting. In parallel, the AI Act introduces a risk classification system, with many AI applications used in CI falling into the high-risk category. This dual alignment implies that AI systems must meet both the cybersecurity standards under NIS2 and the risk management, transparency and oversight requirements under the AI Act. This convergence is central to the EU's strategic approach, ensuring that AI systems used in critical operations are not only effective, but also trustworthy and secure by design.

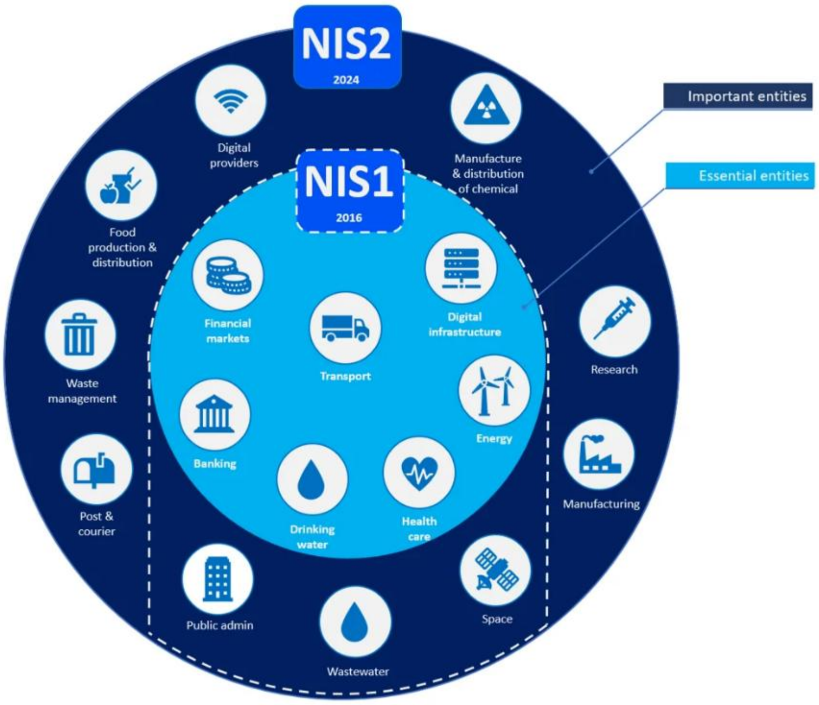


Figure 5: Expansion of scope and sectors under the NIS2 Directive compared to NIS1.

Nevertheless, implementation challenges persist. For example, under NIS2, operators must reconcile the need for real-time emergency response with requirements for incident reporting and forensic evidence collection. Business continuity and disaster recovery planning, both mandatory under NIS2, should also encompass the AI supply chain and third-party risk exposure.

The CER Directive complements NIS2 by focusing on the **physical** and **operational** resilience of critical services as defined under Article 114 of the Treaty of the Functioning of the EU (TFEU). It mandates MSs to identify critical entities and ensures that they conduct thorough RA and adopt mitigation measures. These entities are required to anticipate and report disruptions, and notably, AI tools can play a valuable role in early warning, impact forecasting, and scenario simulation. ATLANTIS activities, such as the Large-Scale Pilot (LSP) 2, have demonstrated how AI can support rapid situational analysis and stakeholder coordination during cross-border emergencies.

The AI Act itself reinforces this structure by requiring clear documentation, transparency over training data, and human oversight for high-risk systems. Although it does not explicitly mandate exhaustive model verification, it calls for risk mitigation, explainability, and lifecycle monitoring. Some recommendations from the ATLANTIS consortium go beyond current regulatory requirements but align with the AI Act's intent, e.g. the need for SBOM, safeguards against model poisoning, and the importance of foundational model traceability.

SUNRISE project findings also emphasized the practical constraints faced by end-users and the importance of maintaining human-in-the-loop mechanisms in high-risk scenarios. However, the AI Act stops short of requiring a strict "human must approve" approach, pointing instead to flexible human oversight based on context and risk.

Ultimately, operationalizing these regulations requires more than compliance. It demands structured cooperation between regulators, CI operators, developers, and sectoral authorities. Effective implementation of NIS2 hinges on timely information sharing and shared threat intelligence, while CER success depends on aligning operational resilience strategies across jurisdictions. RA must be holistic, encompassing internal systems, supply chains and third-party service providers.

As shown in ATLANTIS and SUNRISE, meaningful progress depends on **inclusive** governance. Joint exercises, cross-sector engagement, and the establishment of mechanisms such as an EU AI Resilience Lab (proposed by ATLANTIS) can bridge the gap between policy and practice. Regulatory alignment is no longer optional; it is a shared responsibility that requires continuous engagement and institutional agility.

Policies Recommendations & Strategic Roadmap

In this brief, 'policy recommendations' are understood in a broad sense. They refer not only to formal legislative or regulatory actions, but also to strategic, operational, and funding mechanisms, e.g. those driven by Horizon Europe project consortia that support the implementation of EU policy frameworks for CI resilience.

These recommendations were reviewed and refined with input from the ATLANTIS Policy Taskforce, gathering experts from across the public and private sectors.

Short-Term (0–1 year): Ensure readiness, reinforce trust, and clarify responsibilities

- **Mandate AI system disclosure for CI operators.**
  National competent authorities, in coordination with DG-HOME and the European Union Agency for Cybersecurity (ENISA), should require operators to disclose deployed AI systems and publish RA to foster accountability and prepare for compliance with the AI Act.
- **Deploy minimum SBOM requirements tailored to CI.**
  DG-CONNECT, with support from national Computer Security Incident Response Teams (CERTs), should introduce enforceable SBOM practices for software used in CI, i.e. to enhance supply chain transparency, detect vulnerabilities early and enable coordinated incident response where AI is embedded.
- **Issue joint guidance on high-risk AI under NIS2 and the AI Act.**
  The EU AI Office, ENISA, and MSs digital authorities should co-publish guidance clarifying overlapping obligations for CI operators, supported by concrete use-case examples (e.g. predictive maintenance, automated inspection).
- **Initiate cross-sector trust-building workshops.**

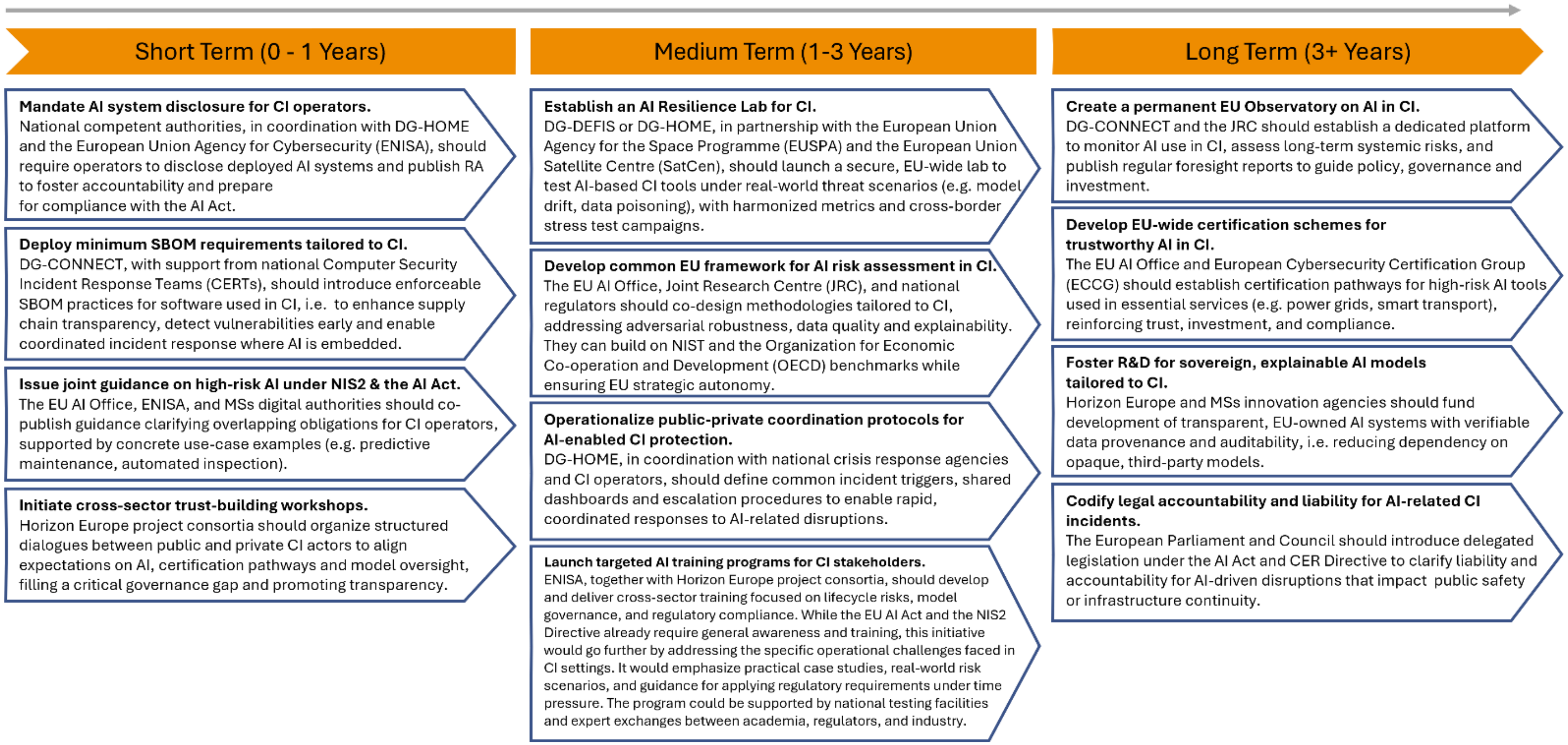
Horizon Europe project consortia should organize structured dialogues between public and private CI actors to align expectations on AI use, certification pathways and model oversight, filling a critical governance gap and promoting transparency.

Medium-Term (1–3 years): Build operational capacity and harmonized risk governance

- **Establish an AI Resilience Lab for CI.**
  DG-DEFIS or DG-HOME, in partnership with the European Union Agency for the Space Programme (EUSPA) and the European Union Satellite Centre (SatCen), should launch a secure, EU-wide lab to test AI-based CI tools under real-world threat scenarios (e.g. model drift, data poisoning), with harmonized metrics and cross-border stress test campaigns.
- **Develop a common EU framework for AI risk assessment in CI.**
  The EU AI Office, Joint Research Centre (JRC), and national regulators should co-design methodologies tailored to CI, addressing adversarial robustness, data quality and explainability. They can build on NIST and the Organization for Economic Co-operation and Development (OECD) benchmarks while ensuring EU strategic autonomy.
- **Operationalize public-private coordination protocols for AI-enabled CI protection.**
  DG-HOME, in coordination with national crisis response agencies and CI operators, should define common incident triggers, shared dashboards and escalation procedures to enable rapid, coordinated responses to AI-related disruptions.
- **Launch targeted AI training programs for CI stakeholders.**
  ENISA, together with Horizon Europe project consortia, should develop and deliver cross-sector training focused on lifecycle risks, model governance, and regulatory compliance. While the EU AI Act and the NIS2 Directive already require general awareness and training, this initiative would go further by addressing the specific operational challenges faced in CI settings. It would emphasize practical case studies, real-world risk scenarios, and guidance for applying regulatory requirements under time pressure. The program could be supported by national testing facilities and expert exchanges between academia, regulators, and industry.

Long-term recommendation (3 years +): Future-proof EU AI governance for CI

- **Create a permanent EU Observatory on AI in CI.**
  DG-CONNECT and the JRC should establish a dedicated platform to monitor AI use in CI, assess long-term systemic risks, and publish regular foresight reports to guide policy, governance and investment.
- **Develop EU-wide certification schemes for trustworthy AI in CI.**
  The EU AI Office and European Cybersecurity Certification Group (ECCG) should establish certification pathways for high-risk AI tools used in essential services (e.g. power grids, smart transport), reinforcing trust, investment, and compliance.
- **Foster R&D for sovereign, explainable AI models tailored to CI.**
  Horizon Europe and MSs innovation agencies should fund development of transparent, EU-owned AI systems with verifiable data provenance and auditability, i.e. reducing dependency on opaque, third-party models.
- **Codify legal accountability and liability for AI-related CI incidents.**
  The European Parliament and Council should introduce delegated legislation under the AI Act and CER Directive to clarify liability and accountability for AI-driven disruptions that impact public safety or infrastructure continuity.



Short Term (0 - 1 Years)

Mandate AI system disclosure for CI operators.
National competent authorities, in coordination with DG-HOME and the European Union Agency for Cybersecurity (ENISA), should require operators to disclose deployed AI systems and publish RA to foster accountability and prepare for compliance with the AI Act.

Deploy minimum SBOM requirements tailored to CI.
DG-CONNECT, with support from national Computer Security Incident Response Teams (CERTs), should introduce enforceable SBOM practices for software used in CI, i.e. to enhance supply chain transparency, detect vulnerabilities early and enable coordinated incident response where AI is embedded.

Issue joint guidance on high-risk AI under NIS2 & the AI Act.
The EU AI Office, ENISA, and MSs digital authorities should co-publish guidance clarifying overlapping obligations for CI operators, supported by concrete use-case examples (e.g. predictive maintenance, automated inspection).

Initiate cross-sector trust-building workshops.
Horizon Europe project consortia should organize structured dialogues between public and private CI actors to align expectations on AI, certification pathways and model oversight, filling a critical governance gap and promoting transparency.

Medium Term (1-3 Years)

Establish an AI Resilience Lab for CI.
DG-DEFIS or DG-HOME, in partnership with the European Union Agency for the Space Programme (EUSPA) and the European Union Satellite Centre (SatCen), should launch a secure, EU-wide lab to test AI-based CI tools under real-world threat scenarios (e.g. model drift, data poisoning), with harmonized metrics and cross-border stress test campaigns.

Develop common EU framework for AI risk assessment in CI.
The EU AI Office, Joint Research Centre (JRC), and national regulators should co-design methodologies tailored to CI, addressing adversarial robustness, data quality and explainability. They can build on NIST and the Organization for Economic Co-operation and Development (OECD) benchmarks while ensuring EU strategic autonomy.

Operationalize public-private coordination protocols for AI-enabled CI protection.
DG-HOME, in coordination with national crisis response agencies and CI operators, should define common incident triggers, shared dashboards and escalation procedures to enable rapid, coordinated responses to AI-related disruptions.

Launch targeted AI training programs for CI stakeholders.
ENISA, together with Horizon Europe project consortia, should develop and deliver cross-sector training focused on lifecycle risks, model governance, and regulatory compliance. While the EU AI Act and the NIS2 Directive already require general awareness and training, this initiative would go further by addressing the specific operational challenges faced in CI settings. It would emphasize practical case studies, real-world risk scenarios, and guidance for applying regulatory requirements under time pressure. The program could be supported by national testing facilities and expert exchanges between academia, regulators, and industry.

Long Term (3+ Years)

Create a permanent EU Observatory on AI in CI.
DG-CONNECT and the JRC should establish a dedicated platform to monitor AI use in CI, assess long-term systemic risks, and publish regular foresight reports to guide policy, governance and investment.

Develop EU-wide certification schemes for trustworthy AI in CI.
The EU AI Office and European Cybersecurity Certification Group (ECCG) should establish certification pathways for high-risk AI tools used in essential services (e.g. power grids, smart transport), reinforcing trust, investment, and compliance.

Foster R&D for sovereign, explainable AI models tailored to CI.
Horizon Europe and MSs innovation agencies should fund development of transparent, EU-owned AI systems with verifiable data provenance and auditability, i.e. reducing dependency on opaque, third-party models.

Codify legal accountability and liability for AI-related CI incidents.
The European Parliament and Council should introduce delegated legislation under the AI Act and CER Directive to clarify liability and accountability for AI-driven disruptions that impact public safety or infrastructure continuity.

Figure 6: *Strategic roadmap for policy recommendations*

Conclusions

The rapid integration of AI-based technologies into Europe's CI presents an urgent governance challenge. As this policy brief outlines, AI systems introduce novel vulnerabilities, i.e. when embedded in essential services that depend on scalable cloud environments, interconnected architectures, and increasingly autonomous systems.

While the EU regulatory framework, anchored in the AI Act, NIS2 and the CER Directive, lays down important foundations, gaps remain in how vulnerabilities are identified, mitigated, and communicated across sectors. Without swift and coordinated action, these weaknesses risk being exploited, amplifying the threat of cascading disruptions across interdependent networks.

These insights, developed through ATLANTIS, underline the urgency of implementing a practical, multi-tiered governance framework for safe and resilient AI deployment in CI. Given the scale, speed and complexity of AI adoption in CI, immediate protective measures must be implemented before systemic vulnerabilities are weaponized.

We therefore call for a timely and coordinated rollout of a **three-tiered policy framework:**

- **Short-term (0–1 year):** Mandate transparency through AI system disclosures and SBOM requirements tailored to CI environments.
- **Medium-term (1–3 years):** Establish an EU AI Resilience Lab to stress-test models in high-risk CI use cases under real-world threat scenarios.
- **Long-term (3+ years):** Implement adaptive oversight via a permanent EU Observatory on AI in CI, complemented by EU-wide certification schemes for trustworthy AI tools.

Effective implementation will require coordinated efforts across all stakeholders:

- **Policymakers** must prioritize regulatory harmonization and provide sector-specific guidance.
- **CI operators** should actively participate in RA frameworks and meet disclosure obligations.
- **Technology developers** must embed trustworthiness and certification into the design of AI supply chains.
- **Horizon Europe projects and the EU AI Office** offer immediate opportunities to pilot and scale these approaches.

The continued integration of AI into Europe's CI must not outpace our capacity to govern it. Only a coordinated framework for risk management, technical safeguards and strategic foresight can ensure AI enhances the resilience and reliability of essential services, rather than threatening them.

References & Sources

[1] European Commission (2025) The Artificial Intelligence Act – Legal Framework, https://artificialintelligenceact.eu

[2] European Commission (2024) Regulatory Framework for Artificial Intelligence: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[3] European Commission (2023) Europe Fit for the Digital Age: Excellence and Trust in Artificial Intelligence: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en

[4] European Commission (2024) Cyber Resilience Act: https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

[5] European Union Agency for Cybersecurity (ENISA). (2023). Software Bill of Materials (SBOM) and Software Supply Chain Security: https://www.cisa.gov/sbom

[6] National Institute of Standards and Technology (NIST) (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce:
https://www.nist.gov/itl/ai-risk-management-framework
https://doi.org/10.6028/NIST.AI.100-1

[7] Organization for Economic Co-operation and Development (OECD) (n.d.). AI Risks and Incidents, https://www.oecd.org/en/topics/ai-risks-and-incidents.html

[8] OECD. (2025). Towards a Common Reporting Framework for AI Incidents: https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/towards-a-common-reporting-framework-for-ai-incidents_8c488fdb/f326d4ac-en.pdf

[9] AI-Watch / European Commission (n.d.) Trustworthy AI: https://ai-watch.ec.europa.eu/topics/trustworthy-ai_en

[10] AI Regulation (n.d.) AI Regulatory Pyramid Visualisation: https://ai-regulation.com/visualisation-pyramid/

[11] European Commission (2025) AI Factories Initiative: https://digital-strategy.ec.europa.eu/en/policies/ai-factories

[12] Verve Industrial (2023) NIS2 Cybersecurity Directive for EU: Implications for Critical Infrastructure: https://verveindustrial.com/resources/whitepaper/nis2-cybersecurity-directive-for-eu/

[13] BSI Group (2024) The EU AI Act and Its Interactions with Cybersecurity Legislation: https://www.bsigroup.com/en-GB/insights-and-media/insights/blogs/the-eu-ai-act-and-its-interactions-with-cybersecurity-legislation/

[14] Argyroudis, S. A., Mitoulis, S. A., Chatzi, E., Baker, J. W., Brilakis, I., Gkoumas, K., Vousdoukas, M., Hynes, W., Carluccio, S., Keou, O., Frangopol, D. M., & Linkov, I. (2022). Digital technologies can enhance climate resilience of critical infrastructure. Climate Risk Management, 35: https://doi.org/10.1016/j.crm.2022.100441

How to cite this brief

ATLANTIS Consortium (2025). *Securing Europe Critical Infrastructure: Aligning AI innovation with policy and governance for Critical Infrastructure resilience. ATLANTIS Policy Brief No. 1 (PIA1). Horizon Europe Project ATLANTIS (Grant Agreement No. 101073909). Brussels.*

Acknowledgements & Contributors

We would like to thank the following reviewers for their valuable comments and feedback. Their input helped strengthen the clarity and coherence of the brief. The final content remains the responsibility of the authors:

Jolanda MODIC	ICS and ATLANTIS Work Package Leader
Marco GERCKE	CRI
Martina FRANCIOSI	CRI

This policy brief was coordinated by **Thomas SELEGNY** (RESALLIENCE), Policy manager of the ATLANTIS project, with contributions of:

David BAKER	UNDRR
Marti FABREGAT	CAIXA Bank
Abla EDJOSSAN-SOSSOU	RESALLIENCE
Eftichia GEORGIOU	KEMEA
Gregor KOVAC	SB-CELJE
Aleksandrina MAVRODIEVA	UNDRR
Aljosa PASIC	EVIDEN and SUNRISE project coordinator
Josip RADMAN	Ministry of Infrastructure of Slovenia
Stefan SCHAUER	ATI
Milan TARMAN	Institute for Corporate Security Studies
Mario TRIVINO	ATOS
Cristian Raul VINTILA	Siemens
Sean TRAVERS	Carr Communications
Rory McGLYNN	Carr Communications

For more information on this brief, contact:

Thomas SELEGNY, thomas.selegny@resallience.com