NIST's generative AI profile: a benchmark for global AI governance and safety

06 June 2024

AUTHORS

Aurora Agostini
Partner

Giorgia Rastrelli
Associate

lexia.it

Introduction

In the rapidly evolving domain of Artificial Intelligence (AI), standardization of practices stands as a cornerstone of sustainable technological advancement. The National Institute of Standards and Technology (NIST), a beacon in the U.S. federal technology sector, has been instrumental in crafting frameworks that not only enhance interoperability among burgeoning technologies but also ensure their safety, reliability, and ethical deployment.

The "Artificial Intelligence Risk Management: Generative AI Profile", together with the "Secure Software Development Practices for Generative AI and Dual-Use Foundation Models", is one of NIST's pivotal initiatives, aimed at establishing a comprehensive set of guidelines that will govern the future integration of AI systems across various sectors.

This article delves deep into the intricacies of this document, exploring its purpose, scope, and the broad-reaching implications it holds for global industries and future technological landscapes.

Purpose and scope

The "Artificial Intelligence Risk Management: Generative AI Profile" document goes beyond offering mere guidance to technology developers and

implementers; it establishes a benchmark for regulatory bodies globally, providing a structured framework to assess AI applications against stringent safety and performance metrics.

In fact, the document aims to promote a unified approach to AI governance, transcending international borders and fostering a global environment where AI technology is both innovative and trustworthy.
The profile's scope is meticulously detailed, covering a wide range of AI implementations across various sectors, including autonomous vehicles, healthcare diagnostics, financial services, and cybersecurity. Each sector faces unique challenges and risks with AI deployment, and the document offers tailored guidelines to meet these specific needs, also ensuring that the standards provided are adaptable and comprehensive, paving the way for a future where AI is integral to societal advancement and individual well-being.

Key Components of the General AI Profile

The "Artificial Intelligence Risk Management: Generative AI Profile" (the "**Generative AI Profile**") is a comprehensive blueprint designed to guide the development, deployment, and management of AI systems while maintaining high standards of safety, efficiency, and ethical responsibility. In this view, the document provides an exhaustive list of risks related to the deployment and diffusion of Generative AI and subsequently a list of actions aimed at mitigating them.

The Generative AI Profile's provision pivot around some key principles applicable both to the technical and ethical dimensions of AI deployment.

- **Safety and reliability:** AI systems shall be safe and reliable under all expected operational conditions. This requisite entails using rigorous testing and validation methodologies to simulate a wide range of scenarios, aiming to identify and mitigate potential failures before they occur. For example, autonomous driving systems are tested in various weather and traffic conditions to ensure consistent performance and safety.
- **Ethical considerations and fairness:** NIST places significant emphasis on embedding ethical decision-making processes within AI systems. This involves designing algorithms that make fairness a core component of the system's operation, minimizing the risk of decisional biases and other ethical issues. In practice, this can mean implementing diverse training datasets to prevent racial or gender bias in AI-driven hiring tools.
- **Transparency and accountability:** AI systems must be transparent in their operations and decisions, allowing users and regulators to understand and trust the technology. This pillar also covers the accountability of AI

developers and operators, ensuring that there are clear lines of responsibility for the outcomes of AI systems. For instance, the decision-making process AI system used for credit scoring, it should be able to explain its decisions using understandable terms so that customers can grasp why certain decisions were made.

- Interoperability and data governance: Ensuring that AI systems can effectively communicate and operate with other systems and infrastructures is crucial for the seamless integration of technologies. This component also covers data governance policies that protect user privacy and ensure the integrity and security of data used by AI systems. Effective data governance helps maintain the confidentiality and availability of sensitive information and personal data, which is especially important in sectors like banking and healthcare.

Technological impact and implementation challenges

The implementation of the NIST's General AI Profile is set to catalyze significant technological advancements by establishing standards for safety, reliability, and ethical practices. However, complying with the high standard set but the General AI Profile comes with some challenges that players cannot ignore, such as the costs and the technological complexity impose to enterprises in order to align with the General AI Profile's standards.

From the regulator's point of view, the main obstacle will be keeping up with this rapidly evolving technology. In fact, the pace at which regulations evolve can lag behind technological advancements, making it difficult for businesses to stay compliant with the most current standards. In order to avoid this misalignment, regulatory bodies should concentrate on developing regulations that can adapt to technological advancements, ensuring that compliance guidelines remain relevant and do not stifle innovation.

Comparative analysis with other global standards

As previously mentioned, NIST's General AI Profile aims at being globally applicable, becoming a milestone in this field not only for United States but for entire world. This global scope represents a nudge for international collaboration and harmonisation of standards. Aligning more closely under the regulatory point of view, can ensure – especially for companies that already operate across borders – smoother operations and integration of AI technologies worldwide, enhancing global commerce and digital diplomacy.

The European Union (EU) employs a meticulous approach to AI regulation with its proposed AI Act, making it one of the most comprehensive frameworks globally. Namely the AI Act adopts a risk-based approach, classifying AI systems according to risk levels and setting stricter requirements for those deemed as "high-risk". Moreover, the EU's regulations focus intensely on consumer protection and transparency, requiring high-risk AI systems to undergo rigorous conformity assessments prior to deployment. By contrast, as discussed above, this NIST approach on Generative AI is much more pragmatic but, on the other hand, it does not ensure flexibility in relation to the evolution of this technology.

A comparative analysis of NIST's standards another international frameworks showcases a spectrum of regulatory philosophies. On this diversity highlights the need for international cooperation to create interoperable standards that support global AI advancements while respecting regional values and legal norms.

Conclusion

The Generative AI Profile Version is a foundational step towards creating a standardized, safe, and ethical framework for AI development and deployment. Establishing international guidelines, not only will foster innovation but also trust in AI technologies. On the other hand, it must be pointed out that global acceptance of NIST's guidelines would give the United States an undoubted advantage: the opportunity to decisively influence developments in an increasingly crucial sector both technologically and economically.

LEXIA'S DATA & TECHNOLOGY INNOVATION TEAM

Aurora Agostini
Partner

Giulietta Minucci
Counsel

Jessica Giussani
Associate

Giorgia Rastrelli
Associate