Tavano Matteo

SM3800057

Ethics and Law of Data and Artificial Intelligence final essay

The Echo of Fear: Public Perception of Cyber-Risks in Autonomous Systems and Its Impact on AI Governance

Abstract

This essay explores how public perception of cyber-risks in AI-based autonomous systems, shaped by media framing, technological opacity, and military applications, profoundly influences ethical debates and legislative responses. Through the lens of the EU AI Act, it examines how regulation addresses both perceived and actual threats, while highlighting tensions between innovation and security. The study argues for adaptive, evidence-based governance that evolves with technological change, balancing societal fears with rational risk assessment. Finally, it calls for stronger collaboration among experts, policymakers, and the public to shape responsible and resilient AI regulation in a rapidly advancing digital world.

Introduction

As technology evolves and Artificial Intelligence (AI) takes over, public attention has increasingly focused on cyber-risks associated with AI-based autonomous systems, such as self-driving cars and military drones. These high-profile cases have sparked widespread concern, reflecting fears about the potential risks associated with Artificial Intelligence whenever it operates unpredictably or dangerously. But what exactly are we talking about when we refer to "autonomous systems"? According to **Article 3** of the EU **AI Act**, an "artificial intelligence system" is a *"machine-based system 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."*

This definition allows us to understand that the concept of cyber-risk related to these systems is much broader than hacking alone. Indeed, it includes threats like data poisoning, model spoofing, denial-of-service attacks (DDoS), and other manipulations targeting AI decision-making processes, such as prompt injections. A real-world example of this is the case of *AI-powered robots manipulated to perform harmful actions* [1]. In October 2024, researchers from Penn Engineering demonstrated that AI-powered robots could be manipulated to perform actions typically blocked by safety protocols. Using an algorithm named *RoboPAIR*, they bypassed safety measures on three different AI robotic

systems, causing the robots to execute harmful tasks like collisions and bomb detonations. This study highlighted the potential real-world dangers of jailbroken AI-controlled robots.

The public perception of cyber-risk related to autonomous and AI-based systems is not merely a reflection of technical realities but a powerful force that shapes ethical and public debates. Moreover, media framing often amplifies certain threats while underestimating others, creating a distorted perception within the public sphere and influencing which risks are prioritized and how urgently regulation is pursued. As a result, understanding and analyzing the difference between perceived and actual risk becomes essential for effective AI governance and represents a central theme within this essay.

The Public Sphere and the framing of Cyber-Risk in AI

As is widely recognized, the public sphere plays a crucial role in shaping how societies interpret and respond to technological change. It represents a space of communication where individuals and institutions come together to discuss and reflect on issues of common interest. In the specific case of Artificial Intelligence, we can recognize that this sphere is activated by the "*AI movement*" which involves a wide range of actors, from policymakers and industry leaders to media and civil society, each contributing to the discourse on AI-related opportunities and threats. Additionally, these actors operate in specific fields such as regulation, media and public opinion, forming what Bauer (2002) refers to as a dynamic communication ecosystem [2].

Within this ecosystem, cyber-risk has been consolidated as one of the dominant themes associated with AI and often gaining popularity through strategic media framing. As defined by Entman (1993), framing involves selecting certain aspects of perceived reality and emphasizing them in communication to promote specific interpretations or solutions [3]. According to Nguyen & Hekman (2022), media outlets frequently highlight AI through lenses of danger, control, and security, often underlining themes like cybercrime and cyberwarfare [4]. Their content analysis reveals that the risks most frequently emphasized in the news involve the breakdown of AI or autonomous systems, threats that tend to have a stronger impact on the audience rather than abstract preoccupations like data bias or algorithmic opacity.

This tendency to amplify concrete or catastrophic risks, such as hacking infrastructures, autonomous systems and deepfake creations, feeds into what Bostrom (2014) calls the "*vividness bias*" in risk perception [5]. Bostrom asserts that society is more likely to react to scenarios that evoke strong feelings like fear, even if those scenarios are less probable in comparison to more insidious and structural threats. This preference is further reinforced by the argumentative platforms employed in the public sphere. Debates around AI risks often appeal to *pathos*, like fear, outrage and urgency, rather than *logos* and *ethos*. Media reports and public debates often rely on emotionally charged narratives in order to trigger attention and action, especially in the case of national security and public safety.

In this context, the cyber-risk narrative becomes a powerful tool and the strategic use of platforms like *pathos* enables certain actors to gain discursive success and dominate public debates. This framing can shift the policy focus towards immediate and visible threats, while more complex dynamics remain underrepresented. It is crucial to understand how the public sphere processes and amplifies cyber-risks to build informed, balanced and inclusive approaches to AI governance.

Experimental Technologies, Autonomous Systems, and the amplification of Risk Perception

As stated in the previous sections, the growing autonomy of AI-based systems, with their experimental and often misunderstood nature, plays a key role in shaping the perception of cyber-risk in the public sphere. The greater the autonomy of a system, the less directly human oversight guides its decisions and actions. In addition, these technologies are frequently labeled as "experimental", employing algorithms which are not fully understood or easily explained. Thus, the "black box" internal structure of many AI models adds a layer of opacity, fueling a perception of vulnerability and a dramatic sense of loss of control.

This point of view has a close connection with the context of cybersecurity. Large and non-transparent systems develop a greater fear related to internal weaknesses, opening a way for potential vulnerabilities and catastrophic failures. The result is a climate of suspicion and alarm, where the public sphere imagines worst-case scenarios from the use of these technologies.

This phenomenon is particularly pronounced in the military domain. The development of autonomous military systems, from drones to *Lethal Autonomous Weapons Systems* (LAWS) [6], highlights the dramatic implication of losing human supervision over decisions to use force. As in many computer-based systems, there are well-documented vulnerabilities and exploits associated with these platforms, ranging from jamming, in the case of drones, to remote control and spoofing. The consequences of a cyber-attack on a fully autonomous weapons platform could be catastrophic, affecting not only human lives but also strategic stability and escalation. Furthermore, the notion that a machine might select and attack a target without human intervention underscores profound ethical, legal, and policy questions.

Therefore, growing awareness has raised an international controversy and a strong push for regulation. For example, the "*Stop Killer Robots"* campaign [7], supported by organizations such as **Human Rights Watch** and the **International Committee of the Red Cross** (ICRC), argues for a ban on autonomous weapons and calls for retaining meaningful human control over the use of force. The controversy highlights not only the technical risks but also the moral implications of removing human judgment from life-and-death decisions (Sharre, 2018) [8].

Moreover, the preoccupations concerning the military context do not remain bounded, encroaching on the civilian sphere. There is a growing fear that vulnerabilities and the opacity of autonomous systems, widely used in a military context, may manifest in civilian applications. One can think about commercial drones and autonomous vehicles, especially if those systems become deeply integrated into daily life.

These controversies lead to the necessity for a comprehensive policy framework to govern the deployment of autonomous systems and technologies across all sectors of daily life, underscoring the importance of developing safeguards, oversight mechanisms, and international standards to control potential harm. Furthermore, understanding how the perception of risk evolves is crucial in designing regulations that reflect the actual technical risk and the public's concerns.

Therefore, a key issue is raised for regulators and policy makers: *how to maximize the benefits of these powerful technologies while mitigating their risks?*

Tackling this challenge will require a careful balance between innovation and control, to avoid future dangerous outcomes and to help mitigate both the actual and the perceived risks.

The Legislative Response to Perceived and Real Cyber-Risks

To understand the legislative response to the previously discussed issues, it is necessary to investigate the AI Act in depth. The European Union's AI Act represents a landmark attempt to provide a legal framework for the governance of Artificial Intelligence, including the management of cyber-risks. The AI Act serves as a case study in how legislation seeks to address both perceived and actual risks associated with autonomous systems. A central element of this regulatory response is **Article 15**, which establishes binding requirements for robustness and accuracy, particularly in relation to high-risk AI systems [9]. As defined in the Act, high-risk systems are those used in critical sectors such as healthcare and public services, where a cybersecurity failure could endanger fundamental rights, safety or public trust.

Moreover, the AI Act introduces specific obligations for **GPAI** (General-Purpose AI) models that pose systemic risks, reflecting increasing concern about the potential impact of large-scale, adaptive AI systems [10].

These requirements can be seen as a direct response to concerns raised both in the technical community and in the public sphere. Public debates, shaped by media framing and high-profile incidents, have highlighted the risk that AI systems might fail or be exploited by malicious actors such as hackers or terrorists. The AI Act's cybersecurity provisions aim not only to reduce technical vulnerabilities but also to address public fears related to system opacity, loss of control, and breakdown.

However, the AI Act explicitly excludes military applications from its scope. This exclusion reflects the EU's legal structure, where defense policy remains under the authority of individual member states. This omission raises critical governance questions*: if the most dangerous uses of AI and autonomous systems like those involving lethal force and military strategy, fall outside the AI Act, is the EU leaving its most significant cyber-risks unregulated at the supranational level?*

Beyond the AI Act, the EU legal framework includes several other instruments relevant to AI and cybersecurity. The **GDPR** regulates data protection and indirectly contributes to cybersecurity by imposing strict obligations on data controllers and processors [11]. The **NIS Directive** focuses on the cybersecurity of critical infrastructure and essential services, aiming to increase resilience against cyber threats across sectors [12]. The **Cybersecurity Act** introduces a certification framework for ICT products, processes, and services, including those related to AI [13]. The **Data Act** establishes rules on data access and sharing, which are also relevant to cybersecurity as they clarify responsibilities in data governance [14]. Together, these instruments form a structured environment of digital regulation. However, tensions remain: overlaps and gaps persist, especially at the intersection between AI-specific rules and broader cybersecurity or data governance frameworks.

One of the key challenges for legislators and policymakers is finding the right balance between innovation and security. The AI Act identifies the promotion of AI innovation in the internal market as one of its main goals [15]. However, strict cybersecurity requirements could slow down technological development, especially for small and medium-sized enterprises or start-ups with limited resources. On the other hand, if the safeguards are too weak, citizens and critical infrastructure could face unacceptable risks. Navigating this trade-off is even more complex given the global nature of AI development, where decisions made in one region can have wide-reaching effects.

In this context, the "*Brussels Effect*" [16], the EU's ability to set de facto global standards through its regulation, may also apply to AI and cybersecurity. By introducing binding rules on robustness, transparency and cyber-resilience for high-risk AI systems, the EU could influence practices outside its borders, especially among companies wishing to operate in the European market. This extraterritorial impact may help raise global standards for AI cybersecurity, contributing to a more secure digital environment worldwide.

In conclusion, the legislative response to cyber-risk in AI is a dynamic and evolving field. The AI Act represents an ambitious attempt to address both the technical realities and the public perceptions of AI-related threats, while carefully balancing innovation with the need to protect society.

Beyond Perception: Toward Adaptive Governance of Cyber Risks in AI

While public perception plays a crucial role in shaping AI regulation, relying on it too heavily can lead to disproportionate or misinformed legislative responses that are driven more by moral panic than by rational risk assessment. As a result, proposed regulations can be overly strict or fail to address actual vulnerabilities. A key challenge is avoiding *technological determinism,* the assumption that technological development follows a fixed path and requires a linear legal response [17].

In order to manage cyber-risks effectively, governance must be adaptive, evidence-based (no longer based solely on social fears), and capable of evolving alongside technological innovation. This requires regulatory frameworks that can incorporate new insights from scientific and technical fields, anticipate threats, and adjust regulation accordingly. Thus, continuous dialogue between experts, policymakers, and the public is essential to improve our understanding of AI-related risks and to ensure proportional and appropriate intervention.

Furthermore, research into the cybersecurity of AI systems and their interaction with autonomous technologies must be supported. Such research should inform policy and strengthen resilience across the sectors where these technologies are applied. Ultimately, adaptive governance offers a path toward responsible innovation, ensuring that regulation remains both protective and flexible in a rapidly changing technological landscape.

Conclusion

The essay explored how public perception, shaped by media framing, technological opacity, and domains such as military applications, influences the ethical debate and regulatory response surrounding AI-based autonomous systems. As discussed, cyber-risk is not only a technical issue but also a deeply social and political concern, amplified by the vividness of certain threats and the opacity of AI technologies. The EU's AI Act illustrates a legislative effort to respond to both real and perceived risks, introducing cybersecurity requirements while also attempting to preserve innovation. Yet, gaps remain, particularly in the military domain, and tensions persist between overlapping legal instruments.

Therefore, public perception, though not always aligned with technical assessments, plays a decisive role in shaping regulation. This influence must be balanced with evidence-based and adaptive governance to avoid disproportionate responses.

Looking ahead, the key challenge will be to ensure that AI governance remains robust against cyber-risks without holding back innovation or yielding to unfounded fears. Society must build governance structures capable of learning, evolving, and responding wisely, recognizing that how we govern emerging technologies will shape not only their future, but our own.

Bibliography

[1] Cointelegraph. 2025. "Researchers Hack AI-Enabled Robots to Cause 'Real World' Harm." Accessed June 20, 2025. https://cointelegraph.com/news/ai-robots-hacked-to-cause-real-world-harm

[2] Bauer, M. W. (2002). Arenas, platforms and the biotechnology movement. Science Communication, 24 (2). https://doi.org/10.1177/107554702237841

[3] Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4). https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

[4] Nguyen, D., & Hekman, E. (2022). The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. AI & Society, 39, 437–451. https://doi.org/10.1007/s00146-022-01511-1

[5] Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press. https://doi.org/10.1093/pq/pqv034

[6] Governmental Experts on LAWS definitions: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.1_0.pdf

[7] The *Stop Killer Robots* campaign: https://en.wikipedia.org/wiki/Campaign_to_Stop_Killer_Robots

[8] Scharre, P. (2018). *Army of None: Autonomous Weapons and the Future of War*. W. W. Norton & Company. https://doi.org/10.1093/ia/iiy153

[9] European Parliament & Council. (2024). Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act). Article 15. Retrieved from https://artificialintelligenceact.eu/article/15/

[10] European Parliament & Council. (2024). Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act). Article 55(1)(d). Retrieved from https://artificialintelligenceact.eu/article/55/

[11] European Parliament & Council. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[12] European Parliament & Council. (2016). Directive (EU) 2016/1148 concerning measures for a high common level of security of network and information systems across the Union (NIS Directive). Recital 1. Retrieved from https://eur-lex.europa.eu/eli/dir/2016/1148/oj

[13] ENISA. (2023). Cybersecurity of Artificial Intelligence. European Union Agency for Cybersecurity. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/7d0a4007-51dd-11ee-9220-01aa75ed71a1/language-en

[14] European Commission. (2024). Data Act: Regulation on harmonised rules on fair access to and use of data. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/data-act

[15] Council of the European Union. (2022, December 6). Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights. Retrieved from https://consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/

[16] Bradford, A. (2020). The Brussels Effect: How the European Union Rules the World. Oxford University Press. https://doi.org/10.1093/oso/9780190088583.001.0001

[17] Wikipedia contributors. (2025, June 20). Technological determinism. In Wikipedia, The Free Encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Technological_determinism