SSA Paper

2024

Author: Ben Mna Yassine
Faculty of International Business and Management

MESTERSEGES INTELLIGENCIA BIZTONSAG AZ EUROPAI UNIOBAN AZ EU AI TÖRVENY KERETEIN BELÜL.

ARTIFICIAL INTELLIGENCE SECURITY IN THE EUROPEAN UNION UNDER THE FRAMEWORK OF THE EU AI ACT

Supervisor: Dr. Andrási Gábor

11/03/2024

Table of Contents

INTRODUCTION

Artificial Intelligence as a new technology has emerged to redefine how activities in various industries are conducted, affect social relations, and complicate legal frameworks. Currently, technology is penetrating sectors such as healthcare, finance, law enforcement, and other systems with several potential prospects to advance the economy. The European Union (EU) understands AI as a driver of efficiency and competitiveness but also as a factor that must address security and ethical implications. AI can unlock economic growth, improve public service, and fuel science and innovation, but it brings substantial risks that need to be well-managed to prevent harm to the public, their data, and associated trust. As the EU strives to capture the AI opportunity, the EU has set an extensive regulatory vision to turn AI into a competitive advantage for Europe, one based on European principles and values and seen through the lens of ethics and morals.

The European Union defines AI in its legislative documents, specifically in the proposed **Artificial Intelligence Act**. In Article 3 of the EU AI Act, AI is defined as software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with, From this definition This act outlines AI systems as software that is developed with a certain level of autonomy to perform tasks that would typically require human intelligence, which includes learning, reasoning, problem-solving, perception, and language understanding.

The EU has set main objectives for the development and implementation of artificial intelligence including the need to support economic growth through technology adaptation and progress, in addition to promoting the application of ethics in AI development. This human orientation is well captured in strategic papers such as the European Strategy for Data or the White Paper on Artificial Intelligence, in which Europe finds its position for a trustworthy and reliable AI ecosystem. In this context, the EU still imagines itself as a global reference point for trustworthy AI, a perspective based on safety, accountability, and transparency. Through this vision, the EU is interested in not only being a technological pioneer but also a model of responsible innovation.

This development is due to the growing development of security risks associated with the increased use of AI across multiple infrastructures like essential public services, industries, and even governments are likely to incorporate the use of AI systems to deliver services, as the use of AI increases so does the possibility of these systems being hacked or used for evil purposes. These threats include deepfake technology and misinformation as well as potential abuses in the military context which threaten privacy, safety, and the democratic system. Algorithms being introduced to high-risk areas such as policing and health care means the potential for mistakes, misuse, or inherent bias in the system which could cause a great deal of harm. The EU has deployed its risky regulatory strategy into the guidelines that ensure ethical, secure, and transparent usage of AI technologies by setting guidelines based on a human-centric approach, respecting all applicable laws and regulations, following ethical principles and values, and being robust in the technical and the policy aspects, all of which are important when It comes to sustaining the public's confidence and its security. Thus, including them in its legal requirements for the promotion of AI, the EU seeks to ensure that advancements in this industry will be both novel and compliant with certain democratic standards and the general needs of society.

Since there is higher sensitivity to the risks that faulty, biased, or malicious AI systems pose, which means that these tools have to be tested to the maximum and checked periodically. In short, the EU understands that AI must be trusted and adopted in several critical applications if the current sufficiently developed infrastructure of artificial intelligence is to be designed for safe and ethical operation.

Another thing to which attention should be paid is that If rigorous supervision of the process is not set, there are high chance that the public will develop a negative attitude towards the implementation of the AI technology as well as the institutions that support it, which in essence will slow down the pace of development and implementation. However, it is also important to note that as nations of the world race to the top to produce some form of AI capability, the EU's strategy of regulation puts it ahead as a model for responsible innovation.

The EU's regulatory approach to AI led to the creation of the Artificial Intelligence Act (EU AI Act), which was introduced by the European Commission in April 2021 and was later enacted by the European Parliament on March 24, 2024. This bill sets out a broad structure of laws to address a multitude of risks posed by AI systems and enables innovation within an appropriate risk limit. The AI Act introduces a risk-based classification system that categorizes

AI applications into four levels: the acceptable level of risk which includes: Unacceptable risk Highly risky Moderate risk minimal risk. Under this classification, the EU can safeguard the extent of regulatory intervention according to the risk posed by each AI application.

While the EU AI Act is comprehensive and encompasses all forms of AI uses, it can fall short of the fast-changing AI technologies environment. AI development is still rapidly progressing, and the requirement to regulate it within inactive legal frameworks is rather problematic; new AI technologies and algorithms, including self-learning ones and complex models, may produce unforeseeable risks that exceed the potential of the existing regulation.

Therefore, this paper aims to evaluate how the EU AI Act constrains and addresses the security risks of AI applications and discuss possible future shortcomings resulting from the constant development of new AI solutions. Through assessing how the proposed Act differentially regulates high and low-risk AI applications, the enforcement measures, and how the proposed Act encourages the disclosure of risks, this research will examine where the EU AI Act is robust and where it can be more adequately designed to counter future security threats. Furthermore, by comparing it with AI security measures in countries such as the United States and China, this work will outline the features that could help the EU improve its regulatory model and possibly strengthen its framework.

Given this structure of the paper, the first chapter outlines a general view of the potential use of AI in the case of the EU, while the second one comprises the view of security threats and concerns associated with the usage of AI. Secondly, it will analyze the current structure of the EU AI Act as well as the strengths and limitations that the work has and how it can be improved. The final chapter will present probable future threats toward AI security that have not been touched upon in previous chapters of the paper and give recommendations on enhancing the current EU AI Act.

As shown below, this thesis will also shed light on how the EU AI Act stands as a security shield for AI and further establish the EU's leadership in establishing standards for amenable AI governance worldwide.

Studying the impact of the EU AI Act can answer several research questions.

1. How effectively does the EU AI Act address the current security risks associated with high-risk AI applications in sectors such as healthcare, finance, and law enforcement?

2. What are the primary challenges the EU AI Act faces in regulating emerging AI technologies, particularly those that involve self-learning algorithms and complex models?

3. How does the EU AI Act's approach to AI security and ethics compare to AI regulatory frameworks in other regions, such as the U.S. and China?

4. What policy adaptations could improve the EU AI Act to better address cross-border AI security threats and maintain consistency with global AI standards?

These research questions can guide the reader through the paper's aims and provide a clear understanding of the aspects of AI security and regulatory challenges that the study will investigate.

Chapter 1: Foundations and Security Concerns in AI Development

This paper has established that every AI application introduced into various sectors creates massive security challenges that need to be met within the framework outlined in the EU AI Act. That is why the current chapter will focus on examining both the theoretical background that contributes to the definition of AI security within the EU context and the historical developments that provided the historical background for the contemporary EU concept of AI security. From the current literature, we are able to understand the widely accepted theories and further establish the relationship between this AI technology and the security issues that the EU AI Act seeks to address. Moreover, this chapter will focus on particular kinds of threats that appear during the stages of both the development and deployment of AI systems and study how these threats are regulated in the framework of the EU AI Act.

1.1 Literature Review and Background

Concerning this part, the research focuses on examining the existing works in the context of AI security, including theoretical frameworks concerning the EU AI Act. Given prior studies, the risks, ethics of AI, and regulatory alternatives can be assessed. The following section provides an understanding of the current literature on AI security in both academic as well as policy domains to identify research gaps and set the scope for analyzing the role of the EU AI Act to fill them. Thus, it is within this context that this paper locates itself within the larger field of AI security studies**.**

AI security research is extensive and includes works that provide preconditions for recognizing threats and potential security implications related to artificial intelligence, as well as those that are relevant to the EU AI Act framework. Experts of security in AI have formulated security from definitions viewpoint, risk management, ethical issues, technical methods, and history. Together these frameworks assist in describing the issues associated with the application of AI within society which we can sense in the EU AI Act.

One influential piece in the field is “The Malicious Use of Artificial Intelligence: Ensuring that mobile devices securely store the NHs’ private information is key; the Brundage et al. (2018) article, Predicting, Preventing, and Mitigating is useful in this endeavor. I shall explore the theme of non-security in this work as the work focuses on the idea of the duality of technologies where beneficial AI technologies can be bent to negative purposes. The authors, continue to posit that AI security needs a risk management strategy because individuals and groups can manipulate algorithms for cybersecurity invasions, fake news, and spying. As a result, the findings in this paper have helped define how decision-makers and scholars see AI threats to affect the security arrangements included in the recently proposed EU AI Act.

It is also becoming apparent that ethical aspects play an important role in the AI security literature including the problems of fairness, transparency, and accountability. The article Weapons of Math Destruction by Cathy O'Neil seeks to present the argument that prejudicial tendencies ingrained in artificial intelligence predictions will result in unfair decisions. It drives high-risk AI systems to be transparent and be held accountable through the provisions stipulated under the EU AI Act. Such a requirement is meant to eliminate bias and make AI technologies run most safely and ethically.

Two significant concepts identified in the literature include a socio-technical approach that adopts the view that AI security is a technology system within a social context. This view is invaluable in explaining how the users engage with the AI systems and how this can be exploited by either intentional or accidental introduction of threats. This brings in the EU AI Act to foster user training and system transparency as a way of overseeing these socio-technical risks.

In practice, the regulation of AI was previously very liberal, and slowly, the change has been observed, and the proposed policy has become more active in recent years, for instance, in the EU Commission's "White Paper on Artificial Intelligence" (2020). This shift is due to the increasing awareness of AI's dangers, as well as the EU AI Act, which outlines guidelines for high-risk applications and holds AI creators and users accountable.

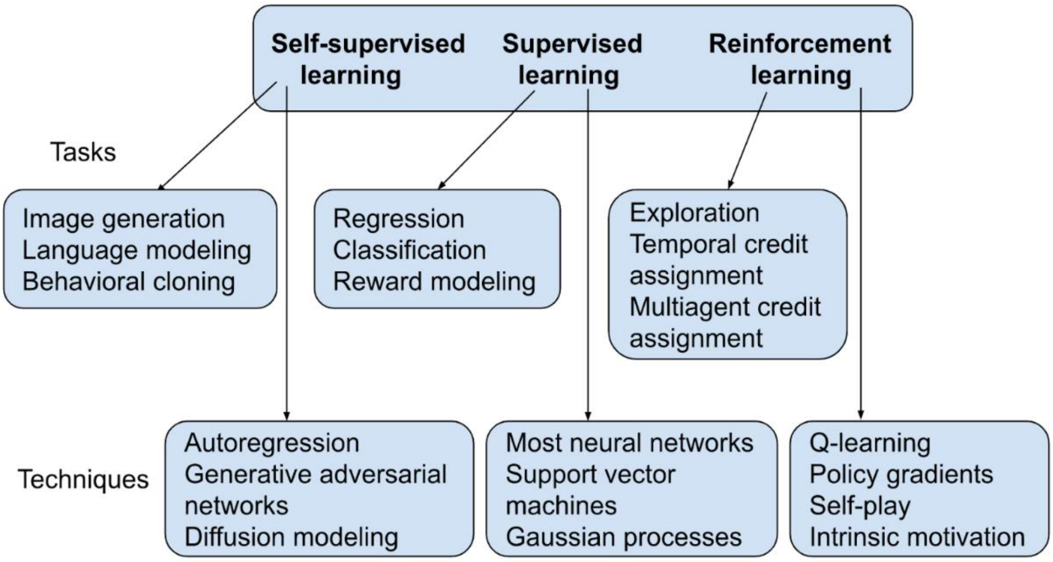
Altogether, the analyzed sources in the field of AI security create a comprehensive knowledge base that underlies and contributes to the EU AI Act's regulation objectives. By incorporating RM paradigms, ethics, socio-technical knowledge, and earlier regulatory changes, the Act is a comprehensive approach to comprehending AI security issues. From this perspective, the EU AI Act can be viewed as legislation alongside the ongoing evolution of her academic and practical discourse on the safe and responsible acceptance of AI in society.

1.2 Security Concerns in AI Development and Deployment

As artificial intelligence (AI) continues to be deployed across industries, it raises deep security considerations, such as a threat to data security, open and exploitable flaws in algorithms, and many ethical concerns. These are addressed by the EU through the EU Artificial Intelligence (AI) Act which seeks to control risks associated with high-risk AI systems. This section analyzes security threats within the AI development and deployment process with references to EU AI Act classifications, the articles that are created to address these threats, and well technical requirements this AI Act contains.

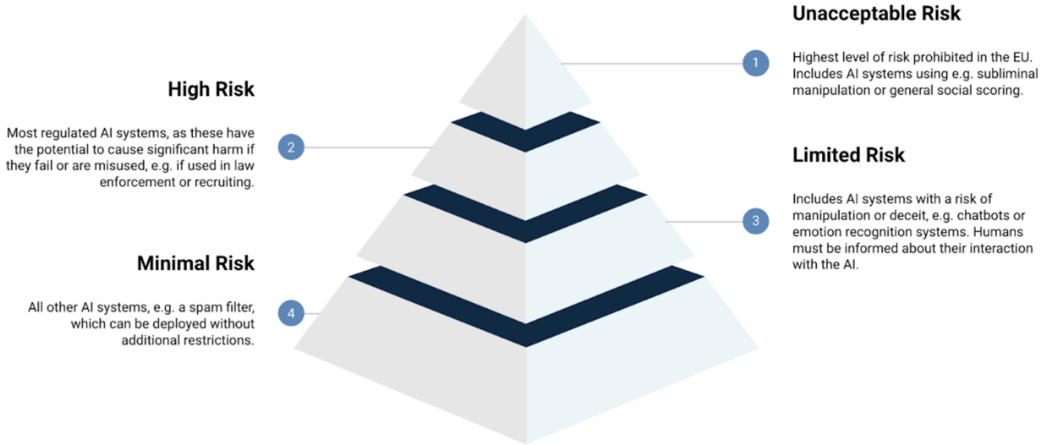
AI systems have already originated several other immense current threats including prejudicial failures, discrimination, social isolation, privacy intrusion, and misinformation. There are also copyright infringements and workers' exploitation in training and deploying the AI systems. Future AI systems might intensify expected catastrophic threats such as bioterrorism or misinformation, misuse of concentrated power, and nuclear and conventional war. We might also hand over the power more or less systematically or unsystematically to the AI systems – or these systems might seize power themselves.

In this image, we can see how AI systems work which not only explains their technicalities but also it raises questions about the possible risks we could face in different steps.



Alignment Forum. (2023). *A Short Introduction to Machine Learning*.

In the same context, the EU categorized the risks in the following way



Highest level of risk prohibited in the EU. Includes AI systems using e.g. subliminal manipulation or general social scoring.

Most regulated AI systems, as these have the potential to cause significant harm if they fail or are misused, e.g. if used in law enforcement or recruiting.

Includes AI systems with a risk of manipulation or deceit, e.g. chatbots or emotion recognition systems. Humans must be informed about their interaction with the AI.

All other AI systems, e.g. a spam filter, which can be deployed without additional restrictions.

(Trail, 2023)

The EU AI Act divides AI applications into four risk categories based on potential harm. At the **Unacceptable Risk** level, AI applications like social scoring and real-time biometric surveillance are banned as they threaten human rights, security, and democratic principles. For example, social scoring systems could unfairly assess individuals based on their behavior or characteristics, while biometric surveillance may lead to privacy loss and unjust monitoring, with biased data potentially influencing machine learning outcomes.

In the **High Risk** category, applications are permitted under strict conditions. This includes critical infrastructure (such as transportation, where AI-related safety incidents have been reported) and education or employment processes, where systems influence lives and thus require rigorous fairness and accuracy testing. Even bias-free AI can exacerbate discrimination in certain contexts; unequal access to AI knowledge and tools, for instance, could increase inequality as wealthier, more educated individuals gain more benefits from advanced AI (AI Safety Fundamentals, 2023). To manage these risks, the EU mandates that high-risk systems undergo risk assessments, ensure transparency, and maintain regular audits. Developers must also design AI to be explainable, robust, and as unbiased as possible.

Limited Risk AI includes systems with moderate risk levels, which require transparency but minimal oversight. Examples are AI chatbots, which must disclose their nature to prevent user misunderstanding, though excessive use could lead to unhealthy relational expectations. Emotion recognition and biometric categorization, allowed here, must notify users of their functions without the high-risk level's comprehensive checks.

Most applications fall under the **Minimal Risk** category, with low-impact uses in everyday tasks that do not need special regulation or transparency. AI for entertainment, such as in video games or productivity tools like grammar checkers and language translators, are considered low-risk, with minimal harm potential. Other examples include algorithms for movie recommendations or ad personalization, which do not generally impact users' rights or safety directly.

The EU AI Act's framework ensures that the regulation of AI applications is proportional to their potential harm. Through this tiered approach, high- and unacceptable-risk applications are carefully monitored while lower-risk applications can innovate with fewer restrictions, balancing public safety with technological progress.

Chapter 2: Evaluation of the EU AI Act's Security Framework

In this chapter, the author focuses on comparing how the EU AI Act ensures the security of AI technologies, relating to different kinds of security. As AI capabilities advance, so do their risks, which can include data risks and bias, as well as risks of misapplication in facilities, services, and systems. To all three, the EU AI Act has provided for extensive regulation to prevent security risks while at the same time encouraging the uptake of trustworthy AI in numerous industries.

2.1 Analysis of the EU AI Act's Security Measures

The core concept of the proposed AI Act is the role that the companies using the software or providing access to it for usage assume. Providers, deployers, importers, and distributors all have standard guidelines for the safe incorporation of AI systems into the market. Providers develop AI systems for the EU market for commercial or non-commercial purposes, and then they must conform to the regulations. Deployers professionally use AI systems with no private use. If the AI originates or is trademarked outside the EU, third-party importers introducing them also have to be from the EU.

The EU Artificial Intelligence Act (AI Act) is one of the most ambitious legal initiatives to prevent the worst negative uses of AI while regulating its development, deployment, and use within the territory of the European Union. The deeper AI becomes incorporated into various fields, the greater the possibility of threats and risks posed by AI systems, which requires different approaches to security threat management. This section examines the security features that have been designed into the EU AI Act, focusing on classification aspects, norms and standards, and enforcement specifications.

One of the foundational elements of the EU AI Act is its risk-based classification system, which categorizes AI systems into four distinct risk levels: unacceptable risk, high risk, limited risk, and minimal risk. This classification allows for tailored regulatory responses that align with the potential impact of each AI application on individuals and society. This risk-based approach enables regulators to focus resources on the most critical areas while encouraging innovation in lower-risk applications.

To curb the risks caused by AI, the EU AI Act sets ambitious regulatory requirements for high-risk AI systems. These obligations are designed to ensure that such systems operate safely and ethically throughout their lifecycle. A risk management system must be established by specialists of high-risk AI, which would encompass all stages of the AI system's lifecycle, including its creation, testing, deployment, and monitoring (Eu parliament, 2023). This is necessary because opportunities should also include an assessment of likely risks in managing data, algorithmic bias, and adverse effects.

Due to the sensitivity of the data used, high-risk systems must follow stringent data management procedures. This means that the training datasets used must be appropriate, diverse, and free of data entry errors (European Commission, 2024). It is crucial for data governance to avoid biases that could lead to discrimination. From a regulatory perspective, providers must maintain up-to-date technical documentation that demonstrates compliance with regulatory requirements. This documentation should be easily retrievable for review by authorities such as Kinstellar (2024). Such disclosure practices are essential for increasing transparency and allowing regulatory authorities to evaluate compliance with safety rules.

According to the Act, providers must monitor and record certain events concerning their AI systems as serious incidents. Any significant loss or breach should be reported to third parties (Vendict, 2024). This requirement promotes accountability and fosters measures to enhance system capability. Currently, the EU AI Act places the principle of accountability at the center of its regulatory measures. Key mechanisms include documentation requirements that mandate the reporting of all artificial intelligence learnings, data sources, and AI decisions made (Vendict, 2024). Such a high level of examination helps providers demonstrate that they operate within ethical and legal parameters.

In risky applications, such as biometric identification or other sensitive uses, providers are required to undertake fundamental rights impact assessments before deployment (2023). These assessments determine the impact on fundamental freedoms, serving as a check and balance against potential abuses. Furthermore, the Act mandates ongoing monitoring requirements for

AI systems that pose risks, promoting changes whenever new threats or weaknesses are identified (Kinstellar, 2024). Continuous supervision is effective in ensuring that existing security measures remain relevant amid the ever-changing challenges in the field of computer science.

Transparency is also an essential component of the EU AI Act's security measures. The Act requires clear communication regarding how AI systems function. Users must be informed when an AI system is being utilized by developers, enabling them to understand how their data is being collected and used, as well as the decisions made by AI instances (European Commission, 2024). High-risk providers are obligated to disclose details about how algorithms operate and under what circumstances (European Parliament, 2023). This disclosure not only enhances accountability but also increases public trust in using AI technologies.

Therefore, the EU AI Act aims to regulate security threats associated with artificial intelligence through a risk-based approach and strict regulation of high-risk AI applications. While the Act promotes the conceptualization, creation, implementation, deployment, and monitoring of AI systems, it provides various methods for managing the identified risks by advocating for accountability and transparency practices in AI system construction. In the coming years, as organizations adjust to these new regulations, understanding these security measures will be crucial for compliance and for promoting AI among the public.

2.2 Comparative Analysis with Other Regions

As AI has progressed and spread to society's tasks, the issue of its regulation has become highly important. Various parts of the world have realized the great significance of AI technologies and have proposed a variety of appropriate regulations considering the tradition and nature of different countries' laws, as well as their culture and economy. This section presents a comparative analysis of AI regulation in three key regions: the European Union (EU), the United States (U.S.), and China.

The EU AI Act remains a landmark attempt to create a rather extensive set of rules concerning AI. Dividing AI apps based on risk, the Act provides the following classifications: hazardous and relevant to safety and important rights. EU unveils strict regulations, especially for high-risk applying AI systems to make sure that these technologies will be created and implemented correctly and without harming the public interest.

On the other hand, the U.S. approach to AI regulation is less rigid and centralized as compared to Other Countries. This is the reason why the regulation often happens on the state level, and there are significant differences in the approach to AI technologies. This principles-based approach advances innovation and endeavors to reduce the regulation burdens to a business entity while it has the problem of a lack of accountability and supervision in high-risk application areas.

On the other hand, China has a developed rather fast satisfactory legal framework for regulating AI with specific concern to recommendation AI and generative AI. The regulations selected for China are targeted at narrow applications and set security for network users. The government has ensured that it enforces the rules to the letter, especially for those operating businesses in the country and from other countries.

This comparative analysis will further also discuss the basic structure of each region's approach to risk categorization, enforcement models, novelty, approach toward data privacy, level of disclosure, and consideration toward public safety. Studying such frameworks helps to understand how distinct world areas adapt to the challenges of governing AI and what further consequences the performed activities may imply for the development of a worldwide framework for regulating AI.

In the following table, we can have a better view of the different approaches of these three powers highlighting their priorities and regulatory work

Aspect	**EU AI Act**	**U.S. Approach**	**China's Regulatory Framework**
Regulatory Framework	Comprehensive and prescriptive	Federated approach with state-specific regulations	Targeted regulations focusing on specific technologies

Key Regulations	AI Act categorizes AI systems by risk (unacceptable, high, limited, minimal)	Executive Order on AI; no comprehensive federal law yet	Interim Measures for Generative AI; PIPL for data protection
Risk Classification	Four-tier risk-based system	No formal risk classification; focus on principles	Conceptual specificity for generative AI technologies
Enforcement Mechanisms	Strong penalties for non-compliance (up to €35 million or 7% of turnover)	Lacks explicit penalties; relies on agency collaboration	Algorithm registry; fines for violations
Focus on Innovation	Balances innovation with safety	Encourages innovation; less regulatory burden	Promotes innovation while ensuring security
Data Privacy Integration	Integrates with GDPR	Varied state-level privacy laws	Governed by Personal Information Protection Law (PIPL)
Scope of Application	This applies to all AI systems deployed in the EU	State-led initiatives; potential for varied applications across sectors	Applies to any generative AI services targeting users in China
Transparency Requirements	Requires transparency for high-risk systems	Encourages voluntary best practices	Requires content labeling and self-assessment for algorithms

Public Safety Considerations	Prohibits certain uses (e.g., social scoring, biometric identification in public spaces)	Focuses on guidelines and best practices	Mandates accuracy and non-discrimination in generated content

The table summarises the key differences between the three approaches and the most peculiar feature of the EU AI Act is its comprehensiveness and prescriptiveness, as well as its risk-based classification system, which is in addition foresighted and concrete, and calls for rather strict compliance measures, especially for the high-risk applications. It focuses on public safety and has a tight collaboration with existing laws regarding data protection, such as GDPR.

The U.S. approach, an executive order, is not built around a federal approach but rather around a federated approach where individual states can pass their legislation. This leads to sometimes a rather loosely connected system of regulation where innovation is fostered without necessarily having strong legal obligations.

Currently, Chinese regulators have paid attention to generative AI technologies through targeted regulations. It focuses on content moderation and how the algorithms work more as it encourages development. The framework also provided heavy penalties for any breach of such rules and required both domestic and overseas service providers who have access to Chinese users.

This table gives a clear perspective on the different approaches used by different regions in the regulation of artificial intelligence and their difference in philosophy and methodology in handling the problem that comes along with these AI technologies.

Chapter 3: Future Challenges and Policy Recommendations

In this chapter, we will discuss the future challenges of artificial intelligence and the initiative whose main reference in Europe is the EU AI Act. As AI technologies develop further new security threats and ethical challenges appear that require constant focus on security regulation and governance. This chapter is divided into two sections: The first of them is to research what possible future threats to security in AI deployment could look like The second is therefore going to be based on policy suggestions that intend to improve the security of AI in the EU. Furthermore, global trend analysis and data from an expert interview will enhance knowledge about these challenges and provide guidelines for effective policy responses.

3.1 Anticipating Future AI Security Challenges

As artificial intelligence (AI) continues to evolve, it brings forth numerous future security challenges that require urgent attention and strategic action. One significant concern is the potential for misuse of AI technologies, particularly by malicious actors. Cybercriminals could harness advanced AI tools to orchestrate more sophisticated cyberattacks, leading to substantial breaches of security and privacy (UK Government, 2023).

This abuse can come in various forms, for instance, the making of deepfake images which have a possibility of twisting the opinion of the people and in the process tearing down the credibility of sources of information. The application of generative RANs introduces critical ethical and security

Furthermore, research shows AI systems are gradually being included in policy and operation of crucial facilities for society such as healthcare, public services, and utilities, where the reliability and robustness of the system assume a higher risk profile (UK Government, 2023). Potential threats to these systems exist and their threats are capable of causing disastrous impacts to the societies within which they exist ranging from transport systems to health services. The report notes that as systems with AI technologies are allowed to become more autonomous problems of who is responsible for what also arise. Some AI algorithms are complex and not transparent enough whose decision-making brings out some harm that was not intended.

To address such issues, the report underlines the value of perfect governance systems in which risk management, compliance, and ethical issues should be considered (UK Government, 2023). The government should therefore ensure international cooperation leading to the development of standards and benchmarks because AI is global and requires a global approach. There is a need for constant assessment and management approaches to keep adapting to ongoing changes in the landscape of AI, to get early warning signs of the risks, and to bring proper preventive measures.

Furthermore, following an expressed call in the report on the proactive study and evaluation of potential threats that AI may pose, will be essential in realizing the potential opportunities of AI without compromising the welfare of society in the UK (UK Government, 2023). Bringing together governments, industry participants and academic institutions, such work will go a long way in unraveling the challenges of AI security and making sure that advancement in this field is not only matched with intelligent solutions to keep people and the community safe.

The issue with AI security is that its development pace is way faster than the production of regulations and legal texts to monitor it, with new AI models created and put out there every day policymakers cannot give instant solutions and policies to regulate it which is why it is a challenging task and with the innovations and improvements of generative AI for example it became difficult to generate laws that are updated to avoid potential risks that are maybe unforeseen or unprecedented.

This approach empowers security teams with the chance to pre-adjudicate so that the system is free of great potential breaches and can be handled by hackers (UpGuard, 2024). For example, AI-based predictive analytical tools can identify unusual user activities and network traffic anomalies that may indicate a cyber-attack and can help organizations make counteractions promptly (World Economic Forum, 2024).

However, as mentioned before the advancement in this field is also fast and brings this new set of problems as well. Hackers are now incorporating AI in their evil deeds to make their plans efficient thus likely to be seen using factors like; synthetic or automated emailed phishing scams that create very realistic emails meant to trick clients (Hornet Security, 2024). Also, adversarial

AI is a real problem of how attackers can take control of machine learning models through adversarial techniques and learn how to camouflage themselves to bypass security measures (eSecurity Planet, 2024). This dynamic nature of cyber threats underlines the importance of the continuous presence of the mind in organizations while employing their security measures.

Also, as the organization starts implementing the AI systems, they bring along with them other vulnerabilities that can be exploited. Thus, data poisoning attacks can cause the inception and manipulation of training datasets thereby producing prejudiced or nonfunctional applications (Forbes, 2024). The reliance on data volume weakens the defense mechanisms through more data exposures opening the undesirable consequences of privacy and security threats. Also, insider threats are a worry; employees may either intentionally or accidentally threaten AI due to automation (Barclay Simpson, 2024).

AI’s duality as a defense enabler and a potential attack surface means that its security cannot be paranoid, but it must be holistic. It is imperative that organizations actively promote the use of AI complemented with preventive strategies on its inherent ethical issues such as privacy and accountability. It finds that by promoting constant surveillance and human intervention, companies can protect their IT properties from new risks in the rapidly evolving information technology environment.

3.2 Policy Recommendations for Enhancing AI Security in the EU

In this section, I draw upon insights from an interview I conducted with Elizabeth James, an expert from the European Network for AI Security (ENAIS). I asked her several questions regarding the effectiveness of the EU AI Act and its ability to manage current risks, as well as potential future challenges and policy recommendations for enhancing AI security in the EU.

James began by discussing the EU AI Act's significance, emphasizing its structured framework for categorizing AI risks into distinct levels—unacceptable, high, limited, and minimal. This classification provides a foundation for robust policy creation while underscoring the importance of ethical standards and human rights. She highlighted that the Act promotes transparency and accountability among diverse stakeholders, including governments and NGOs.

However, she acknowledged that there is room for improvement. Many AI systems rely on historical data that may contain biases, and the field is advancing faster than current research and ethical guidelines can keep up. James suggested that the Act would benefit from standardized regulations across EU member states and closer alignment with international partners to address AI's cross-border impacts. Supporting businesses in innovating within these regulations is crucial, as is refining monitoring and enforcement mechanisms to keep pace with AI's rapid evolution.

When I asked about future challenges in monitoring AI systems, especially those using self-learning or adaptive algorithms, James identified critical issues, including misalignment with human intent and ethics. Continuous adjustments are needed to ensure AI behavior aligns with ethical standards. She also stressed the need for regulatory adaptation, noting that AI technology evolves faster than current laws, as evidenced by cases like Sophia, the humanoid robot granted citizenship in Saudi Arabia, which raises complex questions about autonomy and legal personhood.

James further addressed data privacy and security challenges, highlighting the impact of AI's access to sensitive information on trust and global data governance relations. She also discussed the persistent issues of bias and fairness in AI decision-making, emphasizing the dangers of relying on biased historical data, which can perpetuate discrimination. To tackle these challenges, she advocates for an interdisciplinary approach that ensures AI systems remain ethical, transparent, and aligned with societal values.

Finally, when discussing how the EU can support businesses in adapting to AI security requirements without stifling innovation, James proposed several strategic approaches. She suggested incentivizing compliance through structured programs to encourage organizations to integrate safe AI practices, fostering public trust and market standards. Collaboration between the public and private sectors is also vital for effective policy development. Additionally, prioritizing training and skill development, particularly for startups, will equip businesses to navigate regulatory requirements. Establishing regulatory sandboxes can provide a controlled environment for testing AI systems, enabling companies to address security risks before market deployment. Collectively, these measures aim to create a supportive ecosystem that aligns businesses with regulatory standards while fostering innovation and resilience in an evolving technological landscape.

Conclusion

The emergence of the Program of Artificial Intelligence (AI) opens up truly great prospects as well as poses unparalleled threats in terms of security and ethical values. As described in the previous chapters; to respond to the aforementioned challenges the EU AI Act was developed with the aim to establish a universal regulation of the use of AI that would protect the public's interests and their rights whilst encouraging the further development of the AI market. The ground for this Act has been laid based on a literature review of a large body of data that confirms the necessity to build a strong security framework for the AI integration process. The general public and government need to understand not only that AI systems provide increasing productivity and capabilities but also include potential risks connected with biases, data anonymity, and moral appropriateness.

The Risk Based System of Categorization used in the Act divides AI applications in four different risk levels namely, Unacceptable risk, High Risk, Limited Risk and Minimal Risk which gives a systematic way to approaching all the different implications coming with AI technologies. Besides, it can enable the provision of specific regulatory reactions for each program and guarantee efficient usage of funds for the most significant problem zones. High-risk systems are subjected to strict compliance regimes under the EU AI Act in the form of risk management, data, technical documentation and reporting on incidents. This is especially important as such systems are designed to function in society throughout their life cycle.

Besides, the examination of the EU AI act on security provisions is clear evidence of its concern on transparency and accountability. Moreover, implementing documentation and monitoring in the regular process, the Act aims at strengthening the confidence for the population towards AI technologies and to ensure accountability for any adverse changes keeping developers and deployers responsible for the systems' consequences. But the issue is complex as it comprises of more aspects. The increasing advancement in the fields of AI technologies suggest that the existing legal frameworks must be aligned to the emerging reality. Experts like Elizabeth James from the European Network for AI Security are right in citing the need for constant engagement and partnerships between stakeholders on the issues that surround the best way to monitor AI systems.

Other areas of future development that were mentioned in the third chapter are the necessity to enhance the capability of the regulatory framework to meet future changes and improvements in AI technology, especially in terms of self-learning or adaptive algorithms. AI actions may not reflect the desires, goals, values, and moral compass of humanity; thus we need to make constant tweaks for our AI to be receptive to societal norms. However, matters concerning data privacy, security, and equity in decision-making involving the practice of AI remain significant.

Therefore, AI policy recommendations in the EU must focus on corporate support in the implementation of rules as well as the promotion of innovation. Contemporary methods of compliance encouragement, public-private partnership, and skill prioritization contribute to the sustainable growth of artificial intelligence. Regulatory sandboxes, where the AI systems can operate in a controlled environment to present risks for the intended innovation goal, can be used to advance this cause even with security threats present.

To sum up, one may speak about the EU AI Act as an important step on the way to the correct regulation of such technologies. However, while it advances an excellent start in realizing an AI security framework, continuous assessment and surely reiteration will be essential in responding to new risks and in ensuring that the said framework is relevant to progressing paradigms in information technology. Here, it is crucial to emphasize that only following the principles mentioned above and based on the modern best practices of cooperation and ethical approach, will the stakeholders be able to regain a clearer vision of the further development the EU and further evolution of the AI technologies.

Summary

The EU AI Act sets out a set of rules designed to protect the legal framework for the development and deployment of artificial intelligence. This divides AI applications into four or five risk categories, including no or acceptable risk, high risk, limited risk negligible risk, or minimal risk so it can respond correspondingly. The reality of Artificial Intelligence is considered and regulated in high-risk applications while pointing to transparency, accountability and strict compliance, such issues as bias, protection of personal data, and ethical use. This paper outlines future issues, such as adjustment of regulation to new developments in technology, which require constant discussion among participants. There is also a policy suggestion made to support the new business, partnership, and training for innovation with an adherence to AI security policies.

References

1. AI Safety Fundamentals. (2023). AI risks. Retrieved from https://aisafetyfundamentals.com/blog/ai-risks

2. Bijker, W. E., Hughes, T. P., & Pinch, T. J. (2012). *The social construction of technological systems: New directions in sociology and history of technology*. MIT Press.

3. Brundage, M., Avin, S., Wang, J., Krueger, G., Hadfield, G., Khlaaf, H., ... & Dafoe, A. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Retrieved from https://arxiv.org/abs/1802.07228

4. European Commission. (2020). *White paper on artificial intelligence: A European approach to excellence and trust*. Retrieved from https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

5. European Commission. (2024). *The Artificial Intelligence Act - Regulation (EU) 2024/1689*. Retrieved from https://www.artificial-intelligence-act.com

6. European Parliament. (2023). *EU AI Act: First regulation on artificial intelligence*. Retrieved from https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

7. Jobin, A., Ienca, M., & Andorno, R. (2019). *Artificial intelligence: The global landscape*. *Nature Machine Intelligence*, 1(9), 389-399.

8. Kinstellar. (2024). *The AI Act - EU's first artificial intelligence regulation*. Retrieved from https://www.kinstellar.com/news-and-insights/detail/2577/the-ai-act-eus-first-artificial-intelligence-regulation

9. Moore, A., & Kelsey, M. (2023). *The impact of AI regulation on innovation*. *AI & Society*, 38(2), 300-312. DOI: [10.1007/s00146-023-01234-y](https://doi.org/10.1007/s00146-023-01234-y)

10. Ngo, R. (2023). *A short introduction to machine learning*. AI Alignment Forum. Available at: https://www.alignmentforum.org/posts/qE73pqxAZmeACsAdF/a-short-introduction-to-machine-learning

11. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

12. Trail. (2023). EU AI Act: How risk is classified. Trail ML Blog. Retrieved from https://www.trail-ml.com/blog/eu-ai-act-how-risk-is-classified

13. UK Government. (2023). *Future risks of frontier AI: Annex A*. Retrieved from https://assets.publishing.service.gov.uk/media/653bc393d10f3500139a6ac5/future-risks-of-frontier-ai-annex-a.pdf

14. Vendict. (2024). *The impact of EU AI Act on cybersecurity businesses*. Retrieved from https://vendict.com/blog/the-impact-of-the-eu-ai-act-on-cybersecurity-businesses