Vilnius University Faculty of Law

Lina Sokol

II study year, International and European Law Programme Student

Master's Thesis

Legal Aspects of Application of Artificial Intelligence in National Security

Dirbtinio intelekto taikymo nacionalinio saugumo srityje teisiniai aspektai

Supervisor: Dr Asist. Victor Terekhov

Reviewer: Prof. Dr. Tomas Davulis

Vilnius

2024

ABSTRACT AND KEYWORDS

Abstract. This work analyses the legal framework and challenges associated with the application of artificial intelligence in the national security domain, with a particular focus on European Union legislation. The European Union, as a supranational organization, has adopted a human-centric approach to artificial intelligence governance, which influences national security operations through its supranational security framework. Nonetheless, the European Union cannot deprive national security institutions of artificial intelligence capabilities essential for achieving informational superiority, despite the inherent risks. Doing so could undermine their ability to address internal and external threats effectively, leaving member states vulnerable in an increasingly complex security landscape.

Keywords: artificial intelligence, national security, supranational security, decision-making, informational superiority.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AI Act	Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)
EU	European Union
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
NIS2	Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive)
US	United States of America

INTRODUCTION

The relevance of the Master's Thesis. Advancements in AI are already perceived as one of the pillars in global leadership within international community. As a fundamental technology, AI has capacity to further transform innovations and boost industries. States introduce AI based national strategies and sectoral policies, invest to build robust and trustworthy AI ecosystems. Adoption of the EU AI Act is one of the most significant steps to influence global regulatory approaches that other nations may follow.

Institutional and research capacities are created to analyse possible impacts and outcome of application of AI, and to monitor the regulatory compliance. New international standards are under development. Nonetheless, it remains uncertain which entities will secure credit and international trust as the cyber domain transcends jurisdictional boundaries and cannot be ascribed any single country, region, regulatory framework or stable economic environment. The application of AI and assessment of its capabilities remains in progress with numerous challenges that prevent experts from providing clear assessments of measures implemented today.

Despite of the challenges and associated risks, AI is also integrated to National security strategies and policies as a tool to protect national sovereignty, ensure economic stability and enhance military capabilities. AI applications have the potential to transform traditional national security practices by increasing capabilities in intelligence, threat detection, cybersecurity, resilience and, most critically – expeditious decision making.

Unprovoked Russia's aggression increased tensions in the whole region and especially among neighbouring countries. This aggression exceeds conventional warfare boundaries into cyber domain – cyberattacks, disinformation campaigns, and the use of proxy forces. Such methods can directly influence the stability of a state, undermine their sovereignty or even threaten territorial integrity. Cyber domain lacks formal boundaries enabling hostile state or non-state actors to exert influence without physical presence.

This tension and the increasing threat of expansion of the aggression poses significant risks to democratic societies. Populations may experience insecurity of their future, making them more susceptible to propaganda or to misinformation. Creating such societal conditions hostile state or non-state actors can manipulate public opinion with relative ease and impede workflow of national institutions. A pertinent example is Cambridge Analytica scandal where unauthorised collection of data of Facebook users followed in creation of their detailed psychological portraits and subsequently used to influence voter behaviour in political campaigns, including the 2016 U. S. obstruct Presidential election and the UK

Brexit referendum. Such adversarial actions erode public trust in democratic institutions, their capabilities to uphold democratic values, rights and freedoms within the state.
National security and defence institutions of democratic countries are compelled to employ all lawful measures to provide authorities with timely analytical materials supporting decision-making processes aimed at mitigating threats and averting the escalation of conflicts that could lead to a global war.
The current regulatory framework governing AI application will undoubtfully test the capacity of democratic countries to maintain their legal order and ensure the rule of law through all sectors, including national security.

The aim of this paper is to analyse legal framework and challenges associated with the application of artificial intelligence in the national security domain, with a particular focus on the European Union legislation.
In order to achieve this aim, the following objectives were introduced:

1. To examine the EU's regulatory framework governing AI and compare it to global trends;
2. To explore the intersection of AI, national security, assessing the EU's influence to national security operations;
3. To identify and analyse specific legal challenges arising from AI application that complicate balancing national security objectives, human rights and the rule of law.

Accordingly, **the structure of the Master's Thesis** consists of three main sections. The first section will focus on EU's regulatory framework governing AI, beginning with the analysis of the evolution of the definition of the AI, which delineates the scope regulatory instruments, then the comparative analysis of the current global trends in AI governance, and finishing with the examination of the relevant EU legislature.
The second section is devoted to the examination national security landscape, analysing the intersection of EU Digital policies and national security operations and the aspect of supranational security and proportionality paradigm.
The third section will delineate most challenging aspects of AI and their impact on upholding of the rule of law within democratic societies and operations of national security institutions

Several methods will be employed in this research. First of all, data collection and legal document analysis will be used to gather, categorise and examine online data, literature and

legal documents in order to assess current situation of AI within the European regulatory framework. It will include the review of recent AI Act, other related binding and non-binding instruments like documents of independent High Level Experts Group on Artificial Intelligence, that directly affect evolution of regulatory instruments.

Additionally, a historical comparative analysis will be carried out to analyse the evolution of AI definitions and global trends in the regulatory frameworks governing AI across the EU, US and China to understand the different approaches of leading nations towards balancing AI regulation, ensuring the rule of law and providing national security institutions with necessary tools to safeguard state sovereignty and future stability.

A case law analysis will be conducted to examine judicial decisions concerning the operations of national security institutions, with particular emphasis on the proportionality assessment and the delicate balance between safeguarding the greater good for society and protecting individual rights.

AI itself is not a novel concept. With the appearance of first computers in academia, scholars began to explore ideas about human-like machines (Alan Turing, John McCarthy), capable of learning and solving problems autonomously. They identified general features, that could be attributed to AI systems, including self-learning, language processing and comprehensive output – features that remain relevant today and are included in legal instruments.

The national security nexus within the EU regulatory framework has been analysed through the lenses of cybersecurity, data protection, human rights, counter-terrorism, and the military domain. This paper, however, addresses these aspects as AI risks. Scholars (Reza Montasari) have also explored AI-related risks and challenges to national security, particularly concerning the Internet of Things and big data predictive analysis.

However, the analysis of the AI Act impact to national security remains absent. As the recent regulatory instrument, the AI Act primarily focuses on risk management and assessment of AI based systems within private sector rather than indirect implications for national security and its potential to benefit expeditious decision-making.

I. LEGAL FRAMEWORK GOVERNING ARTIFICIAL INTELLIGENCE: EUROPEAN UNION PERSPECTIVE

The rapid evolution of machine learning, neural networks, natural language processing, generative AI and exponential growth of data centres has transformed both private and public sectors, creating significant economic opportunities while simultaneously raising complex legal and ethical issues. To understand the AI regulatory framework, this section will first examine the evolution of the definition of AI to provide insights into what is regulated and why. This will be followed by an analysis and comparison of global regulatory framework to offer a better perspective on the regulatory approaches adopted by the world's leading nations. The section will be concluded with an in-depth examination of major EU regulations that the author considers have the direct impact on AI application across national sectors, including national security.

1. Evolution of the Definition of Artificial Intelligence

Before any regulation is introduced, definitions must be outlined in a clear and comprehensive manner, as they form the foundation of the regulatory framework, particularly the scope of regulation. Despite the long history of AI research, recent advancements and their significant impact on users' daily lives and market dynamics have urged lawmakers to reassess the necessity of regulating the AI domain.

The concept of AI is widely recognised to have been fist introduce by John McCarthy at Dartmouth Summer Research Project on Artificial Intelligence in 1956 – the term that would come to define the practice of human-like machines (Coursera, 2024). The workshop was based on the conjecture that, "Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves." (Dartmouth Summer Research Project …, 2021) This conference is considered the birth of AI as a research field.

The workshop also established key features of AI that remain relevant today, including the ability to learn, process natural language, provide comprehensive solutions to field-specific problems. Consequentially, computer scientists introduced some groundbreaking developments of AI, such as the first self-learning checkers program by Arthur Samuel in 1959, which could learn from games it had previously played (IBM, The Games that Helped

AI Evolve), or the first chatbot Eliza, developed by Joseph Weizenbaum in 1966, which could repurpose the answers of the users to prompt further conversation (Coursera, 2024).

Another notable figure in the history of AI is Alan Turing, who proposed an empirical test – later known as Turing Test – as a sufficient condition for identifying intelligence in machines that mimics human cognition (Britannica, 2024). The scholar analysed the terms "machine" and "think" (Turing, 1950, p. 1). He suggested that the term machine should encompass three conditions (p. 3): it should include every kind of engineering technique that works, its manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental and finally men, born in a usual way should be excluded. Turing acknowledged the difficulty in framing a definition that meets all three conditions and noted that the concept "thinking machines" is likely inspired by a specific type of machine—the digital computer.

His approach was grounded on an operational perspective, emphasising the ability of a machine to convincingly simulate human responses during a conversation and leading the human interlocutor to believe they are interacting with another person.

However, the advancements in AI in 20th century were subject to critical evaluation. One of the most notable reports, commissioned by the UK Science Research Council, was carried out by Michael James Lighthill under the title "Artificial Intelligence: A General Survey" later known as the "Lighthill Report" (Emanuel, 2024). The report could be considered as the first major attempt to assess the technological level of AI systems and evaluate their potential economic impact. It concluded that no part of the field had the discoveries made so far produced the major impact that was then promised. As a result, it led to the termination of research funding for AI in the majority of universities across UK (Russel, 2020, p. 5), leaving innovations to private sector.

Following the controversial Lighthill's report, a debate of prominent scientists was held at the Royal Institution in London. During the debate John McCarthy described AI as a science that studies problem-solving and goal achieving processes in complex situations (Emanuel, 2024). He also identified for primary challenges within the AI domain: the process of search, internal representation of information within machines, advice-giving and automatic programming. Notably, the definition of AI at this stage remained operational, emphasizing its functional aspects rather than a unified theoretical framework.

It is evident that the early academic analysis of AI has focused on deconstructing the concept into components or fractions that define human cognitive process. This process encompasses discrete features such as reasoning, problem-solving, memory, perception, learning, and language processing. Over decades, computer science experts recreated these

features within machines or programmes. However, this interdisciplinary approach created difficulties for scholars to arrive at a unifying definition. The AI domain encompasses psychology, neuroscience, linguistics, computer science etc. It can be viewed as cross-sectoral application of what is otherwise considered a foundational or general-purpose technology.

General purpose technology (further – GPT) – is described as a single generic technology, recognisable as such over its whole lifetime that initially has much scope for improvement and eventually comes to be widely used, to have many uses and to have many spillover effects (Crafts, 2021, p. 521). This concept was also discussed during the debate, mentioned above and described through the notion of versatility, which means the ability to re-instruct, re-educate rather quickly, easily and conveniently from the point of view of human user for real world situations (Emanuel, 2024).

Additionally, GPTs direct our attention to the boundaries of the definition. Identifying AI systems that qualify as GPT by definition, relies on ex-post success criteria informed by the historical record. GPTs are a subset of ex-post identified growth drivers that are critical to growth (Goldfarb, 2011, p. 822). In essence, GPTs possess the capacity to improve over time through feedbacks from application sectors or end-users, fostering further development and innovation.

Today AI is recognised as GPT primarily due to its transformative potential. It transcends the traditional boundaries of computer sciences and integrates with other disciplines, including psychology, linguistics, healthcare, finances and more, and it changes the way organizations operate, decisions are made, and services are delivered.

AI is also perceived as a critical driver of economic leadership. Nations and corporations invest to build robust and trustworthy AI ecosystems. Studies suggest that AI could contribute to a 3.5% increase in global GDP by 2030, equating to trillions of dollars in economic value (Liu, 2024, p. 1). Taking traditional AI's predictive capabilities further, generative AI tools such as ChatGPT, Co-Pilot, and Midjourney are expected to have substantial impact on economic growth, labour markets, and global trade patterns. The larger the impact to economy growth and markets, the greater the necessity to o implement comprehensive regulatory frameworks to govern these processes.

Every regulatory process begins with defining the underlying concept. As a responsible body, the European Commission has defined AI as a system that displays intelligent behaviour by analysing its environment and taking actions – with some degree of autonomy – to achieve specific goals (Communication from the Commission, 2018, part 1). This definition is intentionally broad, encompassing a wide range of technologies, from robotics

to software applications. The emphasis is placed on the functionality of the technology and its autonomous capabilities, ensuring that the definition aligns with the diverse and evolving nature of AI systems. This definition also underscores the European Commission's shift from robotics, which was initially prioritized as the field for innovations and investments, to a broader and more inclusive approach that encompasses diverse AI technologies.

This definition was further expanded by AI High Level Experts Group (further – HLEG) (European Commission, 2018), established by European Commission to collect comprehensive input and provide informed recommendations on AI-related policies and governance:

"AI systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)."

This definition encompasses the full range of features and aspects of various AI systems, yet it raises more questions than it provides answers, particularly regarding its practical application and scope of regulation.

Another, simplified definition was introduced by the European Commission in its White Paper, describing AI as a collection of technologies that combine data, algorithms, and computing power. This definition holds greater significance from a regulatory perspective, as it directly aligns with the core elements of AI systems, specifically addressing technology, data and informational system (or network). By focusing on these core elements, the EU indirectly addresses key regulatory challenges, such as data protection, algorithmic transparency, system accountability and cybersecurity.

While this simplified definition serves as a practical framework for conceptualizing AI, it cannot be relied upon as a standalone regulatory definition. It lacks the precision and legal

certainty regarding values, required for enforceable legislation. Consequently, during regulatory discussions on AI Act, the EU refined HLEG definition, incorporating key characteristics of AI systems that the EU deems essential to regulate in accordance to EUs values.

Under the AI Act, AI system is defined as 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 (AI Act, Article 3(1)).

This definition explicitly excludes systems, that lack any degree of autonomy and are strictly rule-based, where operations are predefined and executed automatically by natural persons. The varying levels of autonomy present significant challenges for accountability and transparency as they complicate the allocation of responsibility and the traceability of decision-making processes. It also reflects the EU's acknowledgement that advancements in AI technologies have the potential to significantly impact the preservation of human-centric values and, as such, require appropriate regulatory oversight.

A key characteristic distinguishing regulated AI systems is their capability to infer, which refers to the process of generating outputs such as predictions, content, recommendations, or decisions based on input data (AI Act, Preambule, point 12). Consequently, such systems are also classified as generative AI, emphasizing their ability to produce novel outputs. Generative AI systems, such as ChatGPT or MidJourney, have attracted significant governmental attention due to their transformative potential and broad applicability across sectors. Precisely these AI systems are called general purpose technologies.

To conclude, academic definitions play an important role by providing the conceptual foundation for understanding AI. These definitions typically reflect scholars‘ interests, expertise, and creative interpretations, offering insights into AI’s potential capabilities, applications, and limitations. Academic definitions allow researchers and IT specialists exceed the boundaries of AI application and drive innovation, as evidenced by the developments following Dartmouth Summer Research Project on AI in 1956.

In contrast, legal definitions of AI are more generalised or narrow in nature for a few reasons. Firstly, they are designed to create legal consequences when applied, which require precision and clarity to ensure enforceability and consistency within jurisdiction. Legal definitions must delineate the scope of regulation, identifying what constitutes AI. Accordingly, these definitions focus on specific attributes of AI systems like capability to

infer, learn, reason and model combined with a varying degree of autonomy as identified in the AI Act.

Secondly, regulations are generally perceived as potential constraints on innovation as they impose compliance obligations and increase administrative burdens. For instance, the transparency obligation by AI Act delayed Meta's plans to release an advanced version of its AI model in the EU, due to the unpredictable nature of the European regulatory environment (Milmo, 2024). This situation underscores the delicate balance between fostering innovation and ensuring robust regulatory oversight in the AI domain. And this balance between precision and flexibility or aligning global leadership goals with the preservation of human-centric values provide insights into the future of regulatory frameworks across different regions of the world.

2. Global Regulatory Trends in Artificial Intelligence Governance

The year of 2024 marks a significant moment in the legal regulations of the AI as the EU introduced and adopted the EU Act – the first binding regulatory instrument aimed at governing the AI application across various sectors. This structured regulatory framework and related institutional mechanisms were established to enable the EU, its member states and stakeholders to build a trustworthy AI technologies.

The adoption process, which lasted three years, beginning on April, 2021 with the European Commission's proposal to regulate AI within the EU (European Commission. European approach to artificial intelligence), unfolded global developments of regulatory instruments aimed at addressing challenges and opportunities. These instruments align with the primary objectives outlined in national strategies or policies, reflecting the priorities of leading nations. They highlight the critical balancing point between promoting innovation and leadership in AI and safeguarding human-centric values, the rule of law, and democratic principles.

One of the principal documents in the EU's digital strategy was the 2020 Communication from the European Commission titled *Shaping Europe's Digital Future (2020)*, which set the goals for the Europe to become a global leader in digital transformation. Another key Communication, *2030 Digital Compass: The European Way for the Digital Decade* (Communication from the Commission, 2021), addressed the vulnerabilities of Europe's digital ecosystem exposed during the pandemic crisis – increased dependency on critical, often non-EU based, technologies, reliance on a few big tech companies, rise in an influx of counterfeit products and cyber theft, and the impact of disinformation on our democratic

societies. These vulnerabilities clearly fall within the national security domain as they have the potential to impact national economic growth, pose significant cybersecurity risks and influence political landscape of a state.

In response, Commission proposed strategic measures to empower people and businesses to seize a human-centred, sustainable and more prosperous digital future. The President of the European Commission placed particular emphasis on a development of a European Cloud, leadership in ethical artificial intelligence, the establishment of secure digital identities for all citizens, and the significant enhancement of data, supercomputing, and connectivity infrastructures. It is evident that European Commission, in its pursuit of global technological leadership places a strategic emphasis on three key directions – AI, data and networks. These directions already have key regulatory instruments – GDPR, AI Act and NIS2. While GDPR is fully applicable, the EU and its Member States are currently in the preparation phase for the enforcement of the AI Act and NIS2 directive.

It is worth noting that EU already has global leadership in the enforcement of data protection standards and is consistent in implementing other objectives. Although these strategic policies cannot be directly attributed to national security, given that this domain remains the exclusive responsibility of each EU member states, they directly or indirectly influence national security institutions and their strategic priorities through the application of the rule of law, encompassing international agreements, directly applicable EU regulations, and national legislation, which must align with the former two.

However, balancing human-centric European values with aspirations for economic and technological leadership could pose significant challenges particularly in the face of competing markets that impose fewer administrative burdens on stakeholders and businesses. Only time will reveal whether the regulatory instruments introduced, such as the AI Act, will achieve a similar impact in the technological domain as the GDPR.

The most important EU partner as well as global leadership rival is US. Both share democratic values and the rule of law. However, the pursue of global leadership and future strategies present certain legal differences. These differences are clearly visible in AI governance laws and policies.

One of the most significant legal instruments governing the application of AI in the US is Executive Order No. 14110 *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (2023), laying down comprehensive objectives and obligations to govern AI. As a federal level legal instrument, it primarily mandates federal agencies to develop frameworks and guidelines for a responsible and trustworthy AI development and

deployment ensuring the implementation of due diligence measures to mitigate risks and uphold compliance.

Unlike EU AI Act, this Order does not exclude simple AI systems. The definition AI systems is the most general one and refers to any data system, software hardware, application, tool, or utility that operates in the whole or in part using AI (Section 3(c)). This definition effectively covers all technologies related to AI. Nonetheless, the Order primarily directs its obligations towards federal agencies, rather than private sector stakeholders for effective supervision.

The national security interests are particularly extended to dual-use foundational models that pose a serious risk to security, national economic security, national public health or safety, or any combinations of those matters such as by (i) substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear weapons; (ii) enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber-attacks; or (iii) permitting the evasion of human control or oversight through means of deception or obfuscation. (Section 2(k)).

The majority of widely used AI applications today have originated in the United States, making it logical for the U.S. government to address issues related to the misuse of these systems and their associated infrastructure. Ongoing armed conflicts worldwide further underscore the necessity of implementing robust regulatory safeguards to prevent potential misuse and mitigate risks associated with such technologies.

The Order introduces two categories of obligations. The majority are directed at federal government agencies[1] that form part of the national security system and are tasked with fulfilling their due diligence obligations. The most significant and comprehensive section of the order is dedicated to ensuring the safety and security of AI technology (Section 4) and is based on the protection of internal market, domestic AI technologies from both internal and external threats and risks. Beyond administrative obligations, agencies are required:

- classify and monitor dual-use AI systems, especially foundational models deemed to pose national security risks (Section 4.2);

[1] According to 44 U.S.C. 3502(1), term *agency* refers to any executive department, military department, Government corporation, Government controlled corporation, or other establishment in the executive branch of the Government (including the Executive Office of the President), or any independent regulatory agency except the Government Accountability Office, Federal Election Commission, the governments of the District of Columbia and of the territories and possessions of the United States, and their various subdivisions, or Government-owned contractor-operated facilities, including laboratories engaged in national defense research and production activities.

- ensure privacy protection of US citizens (Section 9(b));
- secure AI infrastructure and supply chain from foreign interference or misuse (Section 4.2(c)).

In accordance with objectives, certain obligations are imposed on companies developing foundation AI system (Section 4.2(a)(i)). These include the requirements to provide the Federal Government with information, reports or records regarding training, developing or producing dual-use foundation models, cybersecurity measures, ownership, possession of model's weight, results of performance in red-teams testing.

Additionally, obligations extend to individuals, organisations or entities that acquire, develop or possess a potential large-scale computing cluster. Such entities are required to report any acquisition, development or possession of such clusters, including location and computing power available as such computing capabilities might be misused by hostile states or non- state actors.

To address the prevention of any misuse of US Infrastructure as a Service (*IaaS*) Products by foreign malicious cyber actors, the order imposes obligations to IaaS Products Providers to submit a report on transactions involving foreign person, including verified identities, types of documentation, associated records such as sources of payment, electronic mail address, internet protocol address, types of accounts (Section 4.2(d)). These requirements align closely with the obligations imposed by the EU on supply chain actors, particularly under the AI Act and the NIS2 Directive. Both frameworks emphasize the need for comprehensive oversight against vulnerabilities that could compromise security or operational integrity.

Another significant document in US AI policy formation is *Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfil National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence* (2024) and is binding for national security institutions. This memorandum directly addresses national security objectives such as reduction of the chemical and biological risks that could emerge from AI, because foundational models are perceived as having capabilities to lower barrier for non-experts to perform skilled tasks such as weapon creation.

The Memorandum focuses on few directions and AI governance is one of them. Although governance is sometimes perceived as a potential obstacle to innovation, it is intended to provide clarity and guidelines for national security institutions. All federal agencies are mandated to appoint Chief AI Officer and establish AI Governance Boards to coordinate and govern AI issues (Section 4.2(e)(ii)). Agencies are also required to submit annual

progress report for at least next five years to the President, enabling ongoing assessment and timely adjustments to the framework (Section 6(b)).

Both this memorandum and Executive order 14110 direct the majority of obligations to Federal Government's institutions rather than to private sector, an approach designed to drive innovation by concentrating regulatory efforts on governmental oversight and leadership in AI governance. This strategy underscores the US President's intent to centralize the governance of AI as a transformative national technology and address vulnerabilities that might pose significant risk to national security matters.

Another focus of the memorandum is the development of AI-enabling infrastructure and computational power, including clean energy generation, power transmission lines, and high-capacity fibre data links (Section 3.1(e)(iv)). With the majority of the world's leading AI technologies, such as ChatGPT, Gemini, Copilot, or Claude, being developed in the US, the increasing demand for AI applications has caused increase in energy consumption, for instance, the widely recognized generative AI tool ChatGPT consumes 25 times more energy per query than Google Search (The Brussel Times, 2024). Therefore, US needs to allocate financial resources to infrastructure in order to secure leadership.

The third, and arguably the most critical, focus of the memorandum is resilience building. This encompasses the protection of critical infrastructure, enhancement of cybersecurity, and the implementation of robust risk management strategies to identify and mitigate risks and vulnerabilities within AI systems.

To compare US policies and regulatory instruments to those of the EU, both aim for global AI leadership, however, the policies differ as the US prioritizes national security considerations, emphasizing the development of robust internal AI infrastructure and enhancing resilience within its domestic market. These efforts are underpinned by a commitment to protecting US citizens' privacy and freedom of speech, with AI technologies that are frequently positioned as the central solution to policy challenges.

The EU positions itself as the overarching protector of human rights, adopting a risk management approach to AI assessment. As a peculiarity of its legislative framework, EU institutions are not empowered to directly address national security matters. Nonetheless, such concerns are indirectly tackled through sectoral policies and institutionalisation and will be analysed in the next Section.

China is recognized as one of the world's leading economies, possessing significant capabilities in the production and supply of technologies and electronic goods. China's official policy on AI was announced early in 2017 through the introduction of *New*

Generation Artificial Intelligence Development Plan. AI is regarded as a key driver for global leadership and global competition. The Chinese government has declared that, by 2030, China will become the world's leading AI innovation centre, establishing a strong foundation to rank among the forefront of innovative nations and global economic powerhouses (Section III). To achieve this objective Chinese government has outlined introduced six primary goals:

1. Build an open and collaborative artificial intelligence science and technology innovation system
2. Fostering a high-end and efficient smart economy
3. Building a safe and convenient smart society
4. Strengthening civil-military integration in the field of artificial intelligence
5. Build a ubiquitous, safe and efficient intelligent infrastructure system
6. Forward-looking layout of major scientific and technological projects on a new generation of artificial intelligence

This strategic policy landscape on AI in China has remained consistent and has been complemented by field-specific regulations in response to the transformative advances in AI technologies. For instance, in 2024 China released a new draft regulation on generative AI (Interesse, 2024) addressing critical areas such as securing training data, protecting AI models, and implementing comprehensive security protocols. Similar regulatory efforts have been introduced globally, reflecting a collective recognition of the need for robust governance frameworks to manage AI's rapid evolution and associated risks.

Western democratic countries draw a clear distinction between private and public sectors, prioritising the protection of human rights, including the rights to privacy, business autonomy, provided entities follow the rules. In contrast, China operates within a framework where national security encompasses almost all domains, blurring the lines between public and private spheres.

Under Article 2 of the National Security Law, national security refers to *a status where the national regime, sovereignty, unity and territory integrity, people's welfare, sustainable economic and social development, and other fundamental national interests are immune from danger and external and internal threats and the capability to maintain this status of security* (Congyan, 2017, p. 80). This expansive definition reflects the overarching authority of the state and, in this context, every action of Chinese institutions or stakeholders is perceived as aligned with the interests of the ruling party. National security institutions serve as enforcement mechanisms and, thus, technological expansion, including

advancements in AI and Internet of Things (IoT) devices, is viewed internationally as a potential threat due to unauthorized data collection and intelligence gathering.
Lithuania's national security institutions identify China as a potential threat (National Treat Assessment of the Republic of Lithuania, 2024) due to concerns over cyber espionage and intelligence gathering on internal affairs. US also perceives China as a national security threat, citing its efforts to establish a China-centric digital infrastructure, export industrial overcapacity, expand domestic technology corporations, and access large data repositories (U.S. Department of Defence, 2023, p. 26). China's investments in global digital infrastructure, including next-generation cellular networks, fibre optic cables, undersea cables, and data centres, further contribute to these concerns.

Despite the differences, all nations striving for leadership in AI, remain in an experimental phase, fostering innovation and exploring potential applications of AI systems. To mitigate risks and vulnerabilities, nations introduce sandboxes (EU) or testbeds (US) or foster scientific experimentation (China) enabling the development and evaluation of AI systems under supervised conditions.
Considering the status of the world's leading economies, the US and China are better positioned than the EU to address emerging issues expeditiously. This advantage stems from their centralized governance structures, which enable rapid decision-making and policy implementation. The EU, by contrast, is not a sovereign state but a supranational organization, meaning that every legislative process or issue management is inherently slower and burdened by prolonged negotiations among its Member States. This structural characteristic may pose a significant disadvantage in the rapidly evolving technological landscape, where swift adaptation and decision-making is critical.

3. Principal Regulations of The European Union Pertaining to Artificial Intelligence

AI-based technologies are widely regarded as transformative game-changers, with their effective adoption offering nations the potential to secure global leadership positions. Like other leading powers, the EU aspires to maintain its standing at the forefront of AI innovation. However, this ambition must align with the EU's cornerstone values (European Council, 2024 p. 3): respect for human dignity, freedom, democracy, equality, the rule of law, and respect for human rights. Both general and sector-specific regulatory instruments must adhere to these fundamental principles.

Taking into account the inherent nature of AI, there are three principal regulatory instruments that are intrinsically linked to the technology. These instruments must be applied irrespective of the sector-specific applicability of emerging AI tools, ensuring a cohesive and comprehensive governance framework:

- GDPR, regulating data protection and privacy;
- NIS2 directive, enhancing cybersecurity and resilience for critical infrastructure;
- AI Act, establishing risk-based framework for safe and trustworthy AI.

While additional regulatory instruments, such as the Digital Markets Act or Digital Services Act, impose significant obligations on large online platforms (gatekeepers) to safeguard economic competition and transparency, the aforementioned three can be considered foundational as are particularly critical for the regulation of general-purpose AI models or foundational AI systems, given their broad applicability and transformative potential across multiple sectors.

3.1. GDPR

The global data environment has become increasingly complex, with vast volumes of data being generated, shared, and processed every second. This complexity stems from interconnected network systems such as social networks, the Internet of Things, and large digital platforms, which facilitate cross-border data flows and have the potential to significantly impact national markets and state's political landscape. Data-driven tools can shape public opinion, spread misinformation, or even interfere in electoral processes.

GDPR in the context of AI application in national security is viewed as a supplementary regulation providing regulatory guidelines for the implementation of the rule of law principle and protection of human rights despite the origins of national security institution within EU. Article 6 encompasses all lawful processing that can be applied by national security institutions in their regulatory instruments.

Another significant consideration is that national security institutions cannot operate in isolation to fulfil their functions. These institutions rely on data obtained from various sources, including open-source intelligence (*OSINT*), national registries, and information systems, where the provisions of the GDPR are directly applicable. This interdependence underscores the necessity for national security operations to align with GDPR principles, ensuring lawful data processing and the protection of individuals' rights.

GDPR plays a significant role in addressing abovementioned privacy challenges by providing a legal framework for managing such data complexities. Its relevance to emerging technologies lies in the terms *automated decision making* and *accountability.*

To begin with, personal data is defined very broadly, encompassing any information related to an identifiable individual of any form – digital or physical (Article 4(1)). This definition is further extended to cover the data processing including: the collection, organisation, storage, deletion and (or) usage of data in any way possible (Montasari, 2023, p 22). With the increasing rate of digital data storage, humans alone lack capacity to qualitatively analyse data sets and provide objective analytical outputs.

If national security institutions deploy AI tools, provisions on automated decision making must be met, particularly ensuring the right of an individual to obtain human intervention. (Article 22). This safeguard is incorporated into Directive (EU) 2016/680 (also titled as the Law Enforcement Directive), but only when such decisions produce an adverse legal effect or significantly affect the data subject.

Consequently, in the context of national security, transparency obligations become applicable only if the automated decisions produce a measurable negative impact on the individual. This approach balances the need for national security with the protection of fundamental rights, emphasizing proportionality and accountability in the deployment of AI systems.

Another important aspect of GDPR is the accountability enforcement mechanism. Data controllers must demonstrate compliance with all principles, enshrined in Article 5(1) and are held accountable for any data breach or unlawful data processing. Therefore, the GDPR imposes certain obligations related to implementation of appropriate measures:

- Controllers and processors must keep detailed records of processing activities if the entity employs more than 250 employees or the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects, the processing is not occasional, or the processing includes special categories of data as referred to in Article 9(1) or personal data relating to criminal convictions and offences referred to in Article 10. (Article 30);
- Relevant supervisory authorities must be notified within 72 hours of personal data breaches (Article 33) and to the data subject if data breach is likely to result in a high risk to the rights and freedoms of natural persons (Article 34);
- If core activities involve large-scale processing of sensitive data or systematic monitoring of individuals, organisations (including national security institutions) must appoint data protection officer (Article 37).

These provisions are also incorporated[2] into Directive (EU) 2016/680 which governs data protection in law enforcement contexts.

Under GDPR, the collection, storage, or processing of personal data constitutes a violation of fundamental human rights unless specific lawfulness criteria are met. A notable example occurred in Lithuania, where the media sought access to personal data processed by the Dignitary Protection Service of the Republic of Lithuania (Janonis, 2024).

Although the GDPR is not directly applicable to national security institutions[3], any request by a third party must still comply with certain standards outlined in GDPR. These include specifying the purpose of the requested data and undergoing necessary procedural safeguards, such as a proportionality assessment, to ensure that the request aligns with legal and ethical considerations.

In practice, general regulatory instruments like GDPR often serve as guiding frameworks for such scenarios, even when they are not directly applicable, underscoring their broader influence on data governance and accountability. The similar impact is also anticipated from the NIS2 and AI Act, as these regulatory instruments are designed to establish robust frameworks for cybersecurity and AI governance

3.2. NIS2

The second foundational regulatory instrument for AI application is NIS2, addressing cybersecurity challenges associated with AI systems and their infrastructure. The title itself indicates the existence of a predecessor, the NIS1 Directive, introduced in 201. However, the evolving cybersecurity landscape and the need to address limitations of the original framework necessitated significant updates, especially important to national security institutions.

NIS2, adopted in 2022 is being implemented gradually to allow member states to assess current cybersecurity landscape and to adopt the necessary measures to ensure compliance with mandatory provisions. Member States were required to transpose the directive into national law by October 17, 2024. For instance, new cybersecurity law in Lithuania entered into force on 18 October, 2024.

Since cyber domain is used as a separate war field, Ministry of National Defence of the Republic of Lithuania has been assigned responsibility for the development of

[2] Provisions are regulated respectively in Articles 19, 24, 25, 29, 30, 31, 32;

[3] Dignitary Protection Service of the Republic of Lithuania is one of the institutions of national security system in Lithuania (Law on Basics of National Security of Lithuania, 22^1 section);

cybersecurity policy in Lithuania. This allocation underscores the critical role of cybersecurity as a defensive framework for any nation striving to secure advanced technologies as part of its global leadership strategy.

In comparison to the former Directive, NIS2 holds particularly important provisions for national security framework:

- *Expansion of the Scope of Critical Sectors*. NIS2 incorporates public administration entities providing public services and key sectors, such as energy, transport, healthcare, digital infrastructure, and space (Article 2(2)), as critical sectors for national security consideration due to high risk of possible human rights' infringements;
- *Enhanced Risk Management Assessment*. Additional risk mitigation elements included, such as introduction of incident prevention and detection system and supply chain security (Article 18(1)) in order to early assess risks to systems critical to national defence and security;
- *Incident Reporting and Response*. Article 20(1) mandates that significant incidents must be reported to relevant authorities within 24 hours of detection to receive timely information and thus enable coordinated responses to threats;
- Cross-Border Collaboration. NIS2 provides few cooperation tools – Cooperation Group (Article 14(1)) at strategic EU level and European Cyber Crises Liaison Organisation Network (Article 15(1)) at national operational level;
- Supply Chain Security. Article 18(2) emphasises the importance of to assess and address risks posed by third-party service providers, especially for national security institutions, as vulnerabilities in supply chains can compromise sensitive systems and operations;

In addition to above mentioned provisions, NIS2 also emphasises governance and accountability in cybersecurity measures (Article 21), promotes the adoption of common standards across the EU (Article 16) and enhances public – private cooperation (Article 6) to strengthen overall national cybersecurity resilience. These provisions align with the EU's strategic objectives to achieve high level of security for networks and information systems. Even in the context of national security institutions these measures serve as primary directives to ensure security of cyber domain.

However, accountability provisions are inherently linked to transparency obligations, which can conflict with the operational secrecy typically maintained by national security institutions. These institutions rarely disclose their vulnerabilities or security status due to the potential risk of exploitation.

Collaboration between national security institutions and private entities is often limited to those entities that comply with secrecy and confidentiality obligations, ensuring the protection of sensitive information. Unlike broader public-private partnerships envisioned under NIS2, in national security contexts, information sharing is typically one-directional—from private entities to public authorities—due to the classified nature of the data involved.

In conjunction with the NIS2, additional EU regulatory instruments were required to be transposed, including the EU Regulation (EU) 2019/881 (Cybersecurity Act), which strengthens the mandate of the EU Agency for Cybersecurity (ENISA) (Article 3(1)) and establishes a cybersecurity certification framework for products and services (Article 46). This directly supports NIS2 by ensuring the availability of trusted tools and systems for critical infrastructure and AI systems, thereby reducing vulnerabilities. The Cybersecurity Act also directly relates to AI Act as a mean for the development of the trustworthy AI.

Furthermore, Regulation (EU) 2021/887 establishes the European Cybersecurity Industrial, Technology and Research Competence Centre (Article 3), as the EU's central body for managing cybersecurity-related research, innovation, and deployment of solutions. This aligns with NIS2's goal of improving cybersecurity infrastructure across Member States. To further facilitate cooperation and information sharing, Network of National Coordination Centres must be established (Article 6) as a tool for collective incident response not only among member states, but also between private and public sectors, a key priority emphasized in NIS2 for achieving a harmonized and resilient cybersecurity ecosystem.

Although Article 2(7) of the NIS2 Directive explicitly excludes its applicability to public administration entities that carry out their activities in the areas of national security, public security, defence or law enforcement, including the prevention, investigation, detection and prosecution of criminal offences (Article 2(7)), member states, on the other hand, may choose not to exempt national security institutions and related entities from obligations. This discretionary approach is particularly relevant given the ongoing war in Ukraine and the hybrid warfare methods targeting democratic nations, which have exposed significant cyber infrastructure vulnerabilities among EU Member States.

A notable example illustrating the critical importance of cybersecurity infrastructure is the 2007 cyberattack on Estonia (Buckland, Schreier, Winkler, 2015, p. 26). At that time Estonia was ranked 23rd in e-readiness ratings with the high-level meetings conducted online, electronic voting and electronic banking transactions. However, this digital connectivity also became its greatest vulnerability. In April 2007, a series of coordinated

distributed denial of service (*DDoS*) attacks targeted the Estonian Parliament, ministries, banks, and media institutions. As a result, the websites of the Ministries of Foreign Affairs and Justice had to shut down, while Prime Minister Andrus Ansip's Reform Party website was defaced. The attack also briefly disabled the national emergency telephone number. This incident underscored that the protection of cybersecurity infrastructure is a crucial element of any digitalisation process, particularly for national security institutions and critical public services.

To ensure robust national resilience against hybrid warfare methods, it is imperative that national security institutions adhere to the certain security criteria and obligations outlined in the NIS2 Directive, preventing them from becoming a weak link in the broader cybersecurity framework.

3.3. AI Act

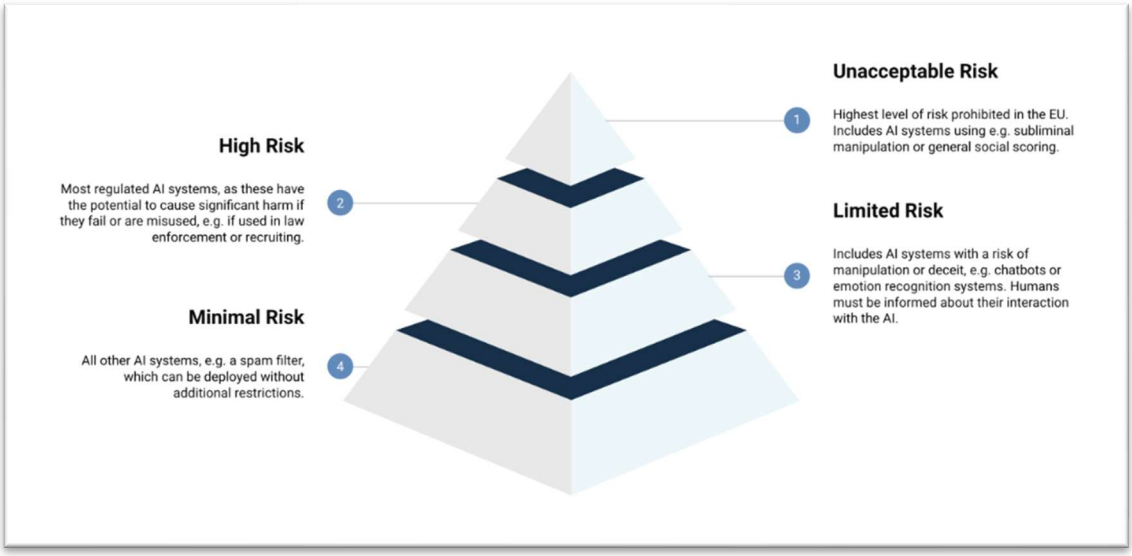
The Strategic Agenda for 2019–2024 primarily focused on addressing pressing challenges such as migration and climate change, rather than prioritizing technological growth. However, the European Commission assumed a proactive role by introducing a key priority for the 2020–2030 decade: A Europe Fit for the Digital Decade. Digital transformation and game-changing technologies are perceived as tools to achieve a prosperous and competitive Europe (European Council, 2024, p. 6).

The rapid evolution and deployment of Artificial Intelligence (AI) technologies have transformed various sectors, creating new economic opportunities while simultaneously raising complex legal and ethical issues. Within the European Union (EU), the application of AI has emerged as both a critical enabler of innovation and a subject of regulatory concern, given its potential implications for privacy, safety, transparency, and fundamental rights. As AI systems become increasingly embedded in society, the necessity for a coherent and comprehensive legal framework governing their development and deployment has become paramount.

The AI Act, adopted in August 2024, establishes harmonized rules across the EU to ensure the development and deployment of human-centric AI technologies. As a transformative technology, AI remains under development, making it challenging to predict its full capabilities and applications. Current expert assessments provide only preliminary estimates of its potential impact. In light of the risks to human rights that AI systems may pose, the European Commission has adopted a risk-based approach that incorporates both current evaluations and forward-looking scenarios to address potential future challenges.

Risk based approach divides AI systems into 4 main categories (See Figure 1).

Figure 1. **Risk based approach of AI Act**



Source: How risk is classified. Available at: <EU AI Act: Risk-Classifications of the AI Regulation>

Prohibited practices, as outlined in Article 5 of the AI Act, are characterized by their potential to cause significant harm to human rights, particularly the rights to human dignity, privacy, freedom of thought, equality, and equal treatment, as well as the equal ability to access services. Although these practices are prohibited, certain exceptions exist for medical or law enforcement purposes, provided they comply with provisions and demonstrably benefit individuals or society despite the inherent risks.

Most regulated AI systems are classified as high-risk under Chapter III of the AI Act. Kalodanis K., Rizomiliotis P. and Anagnostopoulos D. (2024) categorizes these systems into two types: (1) components and products related to safety, and (2) applications of AI in sensitive areas (p. 267). Based on the nature of potential AI system failures, the authors further classify the regulatory requirements into three categories: *Cyber Sec* encompassing all requirements aimed at preventing or mitigating the risk of cyber-attacks, *AI Des* addressing unintentional AI failures, ensuring accuracy, reliability, and robustness of AI systems, and Sys involving logging, documentation, and monitoring obligations to facilitate the adoption, oversight, and accountability of AI systems (p. 269).

Some of these requirements complement related regulatory instruments such as NIS2, which aims to enhance cybersecurity across the European Union. Compliance with these requirements is obligatory and will apply uniformly to all institutions within Member

States. This harmonization reflects the prioritization of security, particularly in the context of crises, where safeguarding critical infrastructure and systems is paramount.

Other requirements related to the use of AI systems in national security remain in a legal grey zone. While the AI Act excludes national security applications from its direct scope, this exception is neither absolute nor entirely unregulated. The rule of law imposes constraints on unsupervised operations by national security institutions, particularly when such operations impact their own nationals.

However, even in cases where individual rights are affected and brought to public attention, the enforcement of individualized redress mechanisms hinges on meeting the threshold of significant harm. Only serious violations trigger legal accountability and remedies, while also allowing national security institutions to operate within the bounds of necessity and proportionality, or so-called grey zone.

Another significant aspect of the AI Act is the institutionalization of AI governance through the establishment of an AI Office, functioning as an EU-level institution (Article 64). On the one hand, this initiative underscores the EU's commitment to centralized governance and harmonized regulatory oversight, promoting uniformity in the enforcement and application of the Act across Member States. This institution is tasked with coordinating enforcement, ensuring compliance with the Act, and providing guidance on the lawful development, deployment, and use of AI systems (European AI Office, 2024).

On the other hand, this institutionalization and regulation may be perceived as constraining the national sovereignty of Member States, limiting their ability to independently address specific issues based on their unique national capabilities. This approach could also be viewed as a potential obstacle to expeditious decision-making in critical situations.

As part of its stated objectives, the AI Act includes provisions to support innovation by promoting the development and deployment of AI technologies in compliance with ethical principles and regulatory requirements. To achieve this, the Act introduces AI regulatory sandboxes (Article 57) and provisions for the testing of high-risk AI systems in real-world conditions (Article 60).

AI regulatory sandboxes are designed to provide a controlled environment that fosters innovation and facilitates the development, training, testing, and validation of innovative AI systems. Member States are mandated to allocate financial resources for the establishment of first national sandbox by 2 August 2026. Furthermore, all other sandboxes must operate under the supervision of competent national authorities, ensuring compliance with the Act's provisions.

Competent authorities are also regarded as safeguards for personal data processing within sandboxes. Pursuant to Article 59 of the AI Act, lawfully collected data may be utilized within sandboxes for the purposes of developing, training, and testing specific AI systems, provided it adheres to established data protection laws and principles. Recognizing the stringent safeguards established under GDPR, the European Commission permits a limited exception for the processing of personal data within the controlled environment.

Another mechanism for fostering innovation is the testing of high-risk AI systems under real-world conditions. The process involves considerable bureaucratic oversight, which is perceived as opposition to innovation, especially for small or medium enterprises with limited finances and human resources. Nonetheless, to permit such testing, the EU mandates the submission of a real-world testing plan, the approval of the market surveillance authority of the relevant Member State, registration in the EU database, the designation of a legal representative by the provider, the informed consent of participants, and the implementation of human oversight.

Considering the role of national security, pursuant to Article 2 of the AI Act, the regulation does not apply to areas outside the scope of Union law and shall not, under any circumstances, affect the competences of Member States concerning national security, regardless of the type of entity entrusted by the Member States with carrying out tasks related to those competences. In practical terms, AI systems that are placed on the market, put into service, or used exclusively for national security purposes fall outside the scope of this regulation.

However, if national security institutions deploy AI systems for purposes beyond national security, such as the employment, they will be required to comply with the provisions of the AI Act. Additionally, national institutions in democratic countries committed to upholding the rule of law must ensure proper documentation of operations to meet accountability requirements. The AI Act and its related instruments provide standardised guidance and frameworks, eliminating the need for institutions to develop individualized documentation plans or systems.

To conclude, national security institutions must leverage opportunities to gather, analyse, and interpret vast amounts of data to make expeditious decisions. AI technologies facilitate these capabilities. However, bureaucratic and institutional obligations may impede decision-making speed or hinder informational superiority of a member state.

GDPR has achieved global recognition for setting a high standard in personal data protection, however, the AI Act’s risk-based approach is already being criticized as a

potential impediment to innovation. Coupled with the EU's extensive regulatory obligations and lengthy legislative procedures, this framework may hinder the Union's ability to secure a leading position in technological transformation in the near future. Nonetheless, the value-based approach may yield significant results over the long term, providing stability, clarity, and a solid foundation for trustworthy and ethical AI development within the region.

II. LEGAL ASPECTS OF NATIONAL SECURITY

"Do not worry about machines taking over the world, do worry about the capacity of either non-state actors or hostile actors to penetrate systems" (Barak Obama interview, 2016 min. 1:35).

National security is a cornerstone of state sovereignty, encompassing the protection of a nation's citizens, critical infrastructure, and core values against internal and external threats. In today's rapidly evolving technological landscape, legal frameworks governing national security must adapt to address emerging challenges such as cybersecurity threats, hybrid warfare tactics, and the deployment of advanced technologies, including AI.

The first part of this section is aimed at analysing national security challenges arising from the cyber domain and the potential integration of technologies. The focus will be on the role of national security in ensuring political independence through informational superiority and expeditious decision-making.

The second part is devoted to the analysis of supranational security emergence within the EU as a result of approach to critical situations posed by external hostile state or non-state actors.

1. Intersection of AI and National Security

Due to the transformative potential of AI and its associated risks to human rights, national security institutions play a crucial role in assessing threats and safeguarding society. However, this role inherently involves a delicate balance between the protection of human rights and the restriction of certain freedoms in the interest of societal security.

The definition and scope of *national security* or *national security interests* can vary significantly across nations, depending on their interpretation of security, current priorities and the nature of internal and external threats they face at any given time. For instance, the environmental protection of international concern was first raised in 1972 in Stockholm with the first UN Conference on the Human Environment (Environment policy: general principles and basic framework, 2024). The priority of deterring Russia's aggression and the subsequent increase in defence budgets are evident among neighbouring countries, reflecting their strategic response to the possibility of the aggression expanding further westward (McGerty, 2024).

Security per se is directly related to strategic policies and objectives. According to Arnold Wolfers (1962, p. 150), security can be understood in two dimensions: objective, which

measures the absence of threats to acquired values, and subjective, which refers to the absence of fear that such values will be attacked. The objective aspect of security is directly related to the responsibilities of national security institutions, particularly, the maintenance of peace as enshrined in the Article 10 of the Convention on the Rights and Duties of States (1933) and the threat assessment to facilitate further strategic objectives for stability and resilience.

For instance, in Lithuanian, intelligence services annually provide the public with a joint non-classified assessment of threats to national security. This practice not only highlights the most pressing threats to the nation, but also intends to educate citizens, encouraging them to remain vigilant against coercive methods employed by hostile states or non-state actors, aimed at disrupting public safety and societal stability.

The subjective dimension of security pertains to society's recognitions of the importance of core values and its determination to safeguard them. The fear aspect reflects nations readiness and technological capabilities to equip military and (or) citizens with necessary means to endure and respond to conflicts of any nature.

The subjective dimension can also be supplemented by Walter Lippman's (1943, p. 51) words – a nation has security when it does not have to sacrifice its legitimate interests to avoid war and is able, if challenged, to maintain them by war. These legitimate interests, encompassing nation's core values, must be perceived by all citizens as having significant importance worth fighting and dying. Otherwise, neither national security institutions, nor defence forces will be able to deter external threats effectively.

Cyber warfare is an integral component of hybrid war and can be examined through certain key characteristics, such as the involvement of state and non-state actors and the associated challenges of non-attribution in both peacetime and during acts of aggression. This form of conflict leverages the ambiguity of cyber operations, complicating the identification of adversaries and the attribution of responsibility under international law.

Under the Charter of the United Nations (1945), States must refrain from the threat or use of force against territorial integrity or political independence of another state (Article 2(4)). The legal framework is straightforward when the aggressor is clearly identifiable, enabling the application of international measures, such as sanctions or collective action under the auspices of the United Nations Security Council. However, the Charter of the United Nations does not explicitly address the use of non-state actors or cyberoperations, which complicates the matter of attribution. Without clear and substantiated evidence linking a state to an act of aggression, the situation cannot be addressed under international law, as

the involvement of another state must be established to justify measures such as the abovementioned self-defence.

Ukraine provides a significant case study for examining offensive cyber operations and the matter of attribution in the context of conventional warfare. On February 24th, 2022, the day of Russia's invasion into Ukraine, a cyberattack disrupted broadband satellite internet access leaving majority of public and private entities without connectivity (Cyber Peace Institute, 2022). This cyber attack was prepared and implemented during peacetime.

The first technical attribution was conducted and publicly disclosed by *SentinelLabs* at the end of March 2022. Experts needed one month to identify certain characteristics linking the cyber operation to Russia and its state actor. A further month was required for public political attribution, presenting references and allegations.

Despite the efforts of experts, establishing ground evidence remained challenging, as the cyber domain exceeds traditional geopolitical boundaries, complicating legal attribution and the application of international law, because Article 2(7) of Charter of the United Nations states, international intervention is prohibited for the matters of domestic jurisdiction.

The problem of attribution significantly hinders the achievement of informational superiority, which is crucial for expeditious decision-making by nations in matters of defence. In conventional warfare, informational superiority has traditionally been associated with defence forces; however, the rapid pace of global digitalization has broadened this objective. Information superiority now extends beyond the defence sector to encompass other critical areas, such as the economy, where access to timely, accurate, and actionable information is essential for safeguarding national interests and ensuring resilience against emerging threats.

In a traditional framework, information superiority comprises two key dimensions: physical and cognitive (Perry, Signori, Boon, 2004, p. xvi). The physical dimension concerns the quality, accuracy, and accessibility of information, while the cognitive dimension relates to situational awareness and the ability to interpret and derive actionable insights. Together data quality and expertise in understanding the data provide timely decisions for national security institutions and governments.

AI provides capabilities to fulfil the requirements of the physical dimension of informational superiority. By properly addressing data bias in algorithms (analysed in Section 3 of this paper), AI can offer national security institutions qualitative analytical insights. However, expertise remains essential to achieve the cognitive dimension of informational superiority. AI technologies lack the interpretative capacity to analyse

individualized situations effectively. As Robert Fein and Bryan Vossekuil (2000) observe, in most protective intelligence cases, based on the gathered information, investigators determine that an individual does not pose a risk to a public figure (p. 55).

The requirement for expertise is directly tied to meaningful human participation in the decision-making process. Processes involving only nominal [4] human participation present the same risks as those completely lacking human involvement (Sancho D., *Algorithms and Law*, 2020, p. 143). AI systems are inherently limited in their ability to engage in the iterative process of legal reasoning, which requires continuous back-and-forth analysis between facts and law, the resolution of contradictory rules, and the management of complex or ambiguous cases. (G'sell F., *Cabridge Handbook of Artificial Intelligence*, 2022, p. 363).

The importance of human oversight is further reinforced by EU legislators. According to Article 14 of the AI Act, human oversight is mandatory for high-risk AI systems. Additionally, when automated decisions produce legal effects or significantly impact an individual's health, safety, or fundamental rights, the affected individual has the right to obtain a reasoned explanation regarding the main elements of the decision (Article 86).

In conclusion, in critical situations, national security institutions are under pressure to deliver timely and accurate predictions, threat assessments, and analysis to enable expeditious decision-making for political leaders in critical situations. Technological capabilities offered by AI are increasingly perceived as superior to traditional intelligence-gathering methods. Considering both dimensions of security – objective (through threat detection, assessment, and mitigation) and subjective (societal trust in security measures) – AI systems play a facilitative role by enhancing operational efficiency and fostering confidence in their capabilities and capabilities of their deployers to mitigate or eliminate threats to society.

In the event of incidents, AI advancements provide technical capabilities for experts to attribute and investigate cyberattacks, gather intelligence, and produce qualitative outputs necessary for informed and prompt decision-making. Nonetheless, such decisions must still rely on human expertise, as AI systems have yet to achieve the capacity to address complex, individualised situations involving multiple and evolving variables. AI systems cannot independently deviate from established parameters, exclude irrelevant variables, or adapt

[4] According to Sancho D., nominal participations refer to participations lacking any real influence on the outcome.

to novel or unforeseen scenarios. Therefore, national security institutions must ensure robust human oversight to ensure informational superiority of a state.

2. Legal Implications of Supranational Security Measures within the European Union

EU is a supranational organisation with clearly divided competences between member states and EU's institutions. Pursuant to Article 4(2) of the Treaty of the EU, national security remains the sole responsibility of each Member State. However, global or regional crises, such as migration crisis, Covid-19 pandemic or Russia's unprovoked aggression against Ukraine, voluntarily or not empowered EU institutions to address national security issues on the European level.

The EU supranational security paradigm was analysed by Ido Sivan-Sevilla (2023). He observes that the nature of transboundary and technology driven security issues lead to surprising and unexplored deployment of supranational policy instruments through the collective involvement and recognised competence of EU-level institutions (p. 1353). The supranational security state is based on three predominant EU policy instruments: export controls over dual-use technologies, network and information security and border security (p. 1354).

Another scholar Federico Casolari (2023) argues that the legal framework covering national security becomes less clear due to the growing marginalisation of the distinction between the concepts of "national security", "internal security" and "public security" (p. 324). The author noted that the European Commission applies to the concept of "public security" activities related to essential interests of member states, which aligns with the Court of Justice of the European Union's (CJEU) interpretation of activities aimed at safeguarding national security—namely, the protection of essential state functions and the fundamental interests of society (p. 325). Moreover, as the notion of "public security" is tied to the concept of European citizenship, it inherently invokes the competence of the European Commission.

Considering scholarly insights and the overarching policies and regulatory frameworks of the EU, supranational security is reinforced through institutionalization and the establishment of EU strategic priorities that address the most critical areas affecting the Union as a whole. The Treaty on EU establishes the EU's seven primary institutions (Article 13), and over the years, numerous specialized agencies have been created to tackle specific challenges. This process of institutionalization inherently expands the competences of the EU.

In accordance with the three primary regulatory instruments analysed in Section One, three executive bodies have been established: the European Data Protection Supervisor, the EU Agency for Cybersecurity (further – ENISA), and the AI Office. Through these centralized supervisory institutions, the concept of supranational security is operationalised and enforced. For instance, the first NIS Directive (2016) expanded ENISA's role by mandating national cybersecurity strategies (Article 1(2)(a)), thereby increasing its scope of involvement. With the introduction of NIS2, the scope of sectors was expanded to include public administration entities of central governments as critical infrastructure elements (Annex I).

The policy expansion is also supported by the Court of Justice of the European Union. As stated in Case C-623/17 (2020), "although it is for member states to define their essential security interests and to adopt appropriate measures to ensure their internal and external security, the mere fact that a national measure has been taken for the purpose of protecting national security cannot render EU law inapplicable and exempt the member states from their obligations to comply with that law".

The same principle applies to the second supranational security enforcement mechanism: the establishment of EU strategic priorities that address critical areas of the Union. The EU's Strategic agenda for 2024-2029 [5] highlights increased engagement in areas traditionally ascribed to national security institutions, including democratic resilience, support for Ukraine, defence readiness and capacity, the defence industry, and domestic crime prevention. These priorities are shaped by the pressing challenges currently confronting the EU, including Russia's war of aggression against Ukraine or the situation in the Middle East. These critical situations clearly impact economic and political stability within states, especially small neighbouring countries with limited resources. Consequently, these countries may voluntarily cede aspects of their exclusive national security prerogatives in favour of regional security and collective prosperity.

A notable example is 2015 migration crisis, during which Member States accepted supranational intervention over national security concerns to share the burden. Regulation (EU) 2016/1624 marked the establishment of the European Border and Coast Guard, tasked with mitigating potential future threats at external borders and ensuring a high level of internal security (Article 1). It also included a proposal by the European Commission (COM(2015) 671 final) to establish a fully operational European Border and Coast Guard

[5] The European Union and Member States have taken bold steps to strengthen the Union's defence readiness and capacity, including increased defence spending.

Agency, evolving from Frontex (p. 3). This agency is designed to support national authorities and, when necessary, substitute national capacities.

Today's pressing challenge is the cyber domain and achieving cyber resilience, which transcends traditional boundaries but directly impacts national security capabilities. Initially, the justification for EU intervention was grounded in economic considerations; however, the rise of cyber-related crimes necessitated action at the EU level (Silvan-Sevila I., 2023, p. 1364). The COVID-19 pandemic and Russia's aggression against Ukraine further underscored the importance of safeguarding the EU's economy against digital threats.

The adoption of NIS2 in 2024 introduced stringent obligations aimed at enhancing the overall level of cyber resilience, protecting critical infrastructure, and enforcing robust security mechanisms for data protection. While the Directive incorporates exclusions for safeguarding national security and their power to safeguard other essential State functions (Article 2(6)), many of its obligations indirectly apply to national security institutions, for example, through the determination and protection of critical infrastructure elements.

Every step towards supranational security is irreversible. If once admitted and regulated, it becomes shared competence with the EU and EU gets power to enforce regulatory requirements concerning national security issues in peaceful periods. This expansion is also visible in case law through the application of principles of the rule of law and proportionality. In particular, the proportionality principle has gained a pivotal role in guiding national authorities and courts (Casolari F. 2023, p. 324).

The concept of proportionality has been extensively analysed by Aharon Barak (2012) through the framework of purposive interpretation. Proportionality is typically described as a criterion, determining the proper relationship between the aims and the means, however, only when the social importance of the benefit in realising the proper purpose is greater that the social importance of preventing the harm caused by limiting the right, can we say that such limitation is proportional (p. 132).

The principle of proportionality is enshrined in Article 5(4) of the Treaty on European Union. It is further elaborated in Protocol No. 2 on the Application of the Principles of Subsidiarity and Proportionality, which mandates that all draft legislative acts must be substantiated in terms of their compliance with the principles of subsidiarity and proportionality (Article 5). In critical situations, necessary and appropriate measures must be implemented to mitigate risks and ensure collective security. However, unlike case law, such measures lack individualisation and, when adopted at the EU level, may inevitably encroach upon the responsibilities traditionally vested in national security institutions.

Addressing crises at the EU level collectively offers significant advantages compared to individual efforts by Member States. The EU has established a comprehensive institutional framework with specialized expertise in critical sectors, coupled with financial resources that exceed the capabilities of any single Member State's budget.

Nonetheless, the centralisation of all resources shifts the informational superiority capabilities from member states to the EU, but does not ensure it due to the decentralised and long decision-making process. Moreover, the application of the proportionality principle further postpones necessary implementation of measures as was seen at the beginning of the Russia's aggression.

Therefore, despite the EU's intrusion into national security scope and intense regulation, national security institutions of member states should preserve their sovereignty in internal and external nationals matters, should allocate necessary resources for technological and expertise capability building for informational superiority and expeditious decision making which would ensure subjective dimension of the notion of security and public support in critical periods.

To conclude, critical situations such as the 2015 migration crisis or Russia's unprovoked aggression against Ukraine have demonstrated a trend where Member States voluntarily cede their sovereign authority to address internal challenges independently. Once regulated at the EU level, such competences must align with EU regulatory provisions, establishing a supranational security framework. This framework permits EU intervention in areas traditionally reserved for national security. As a result, despite national security exclusions in EU regulatory instruments, these provisions are indirectly applicable and must be adopted by national security institutions across the EU.

III. LEGAL ISSUES AND CHALLENGES OF AI APPLICATION IN NATIONAL SECURITY

The integration of AI into national security frameworks offers transformative opportunities. AI technologies hold the potential to significantly enhance intelligence gathering, threat detection, expeditious decision-making, and institutional resilience. However, the full extent of AI's transformative potential remains unexplored. Despite efforts by experts to incorporate safeguards into AI systems, individuals with malicious intent or creative misuse often find ways to alter their primary purposes. This ability, combined with the intentions of hostile state or non-state actors, poses substantial risks not only to individual's rights and freedoms, but also to the stability of political systems and democratic institutions which is the main responsibilities of national security institutions.

This section focuses on the analysis of potential risks and threats posed by AI systems, with particular emphasis on general-purpose AI systems due to their far-reaching impact on society. The analysis is based on the legal frameworks provided by the U.S. Executive Order 14110 and the AI Act, both as binding instruments guiding the governance and regulation of AI technologies.

The table below outlines the AI-related risks identified in the aforementioned legal instruments, providing a clear basis for comparison. As previously noted, the US concentrates the majority of AI governance obligations on federal government institutions rather than the private sector. This approach underscores the government's proactive role in the adoption and oversight of AI technologies, reflecting strategic objectives aimed at safeguarding national interests and ensuring compliance with established legal frameworks.

Table 1. **Comparison of AI risks outlined in EU's AI Act and US Executive order 14110**.

AI Act	**Executive Order 14110**
Risks related to discrimination and bias	Bias and discrimination
Transparency risks	Lack of transparency
Cybersecurity risks	Security threats (infrastructure and cybersecurity)
Risks of misuse in hybrid warfare or crime	Misuse of AI
Data privacy and security risks	Privacy violations
Economic risks	Economic disruption

Threat to human dignity, safety, or fundamental rights	
Accountability risks	
Environmental risks	

Source: Compiled by the author based on EU's AI Act and US Executive order 14110.

The EU, by contrast, has no competence in national security affairs of its member states. As a result, its regulatory approach applies to all entities and is predominantly human-centric, focusing on safeguarding fundamental rights and fostering economic growth as one of its foundational pillars. While this emphasis aligns with the EU's core values, it does not directly correlate with national security interests given that national security institutions are often implicated as major legal violators of human rights. As a result, the EU's approach to AI governance diverges from frameworks that prioritize state security considerations over individual liberties.

Therefore, as major risks outlined by EU and US overlap, further in detail shall be analysed the following risks posed by AI systems to:

- Privacy;
- Bias and discrimination;
- Infrastructure and cybersecurity;
- Transparency and accountability;
- Economic disruption
- Intentional misuse

1. Privacy

Over the past decade, there has been an exponential increase in both the number of data centres and the volume of data generated. Today US is the leading nation in the number of data centres, exceeding other nations by 10 times (see the table…). The proliferation of digital technologies, fuelled by advancements in AI, the Internet of Things, and cloud computing, has led to an unprecedented generation of data. Estimates suggest that by 2025, approximately 181 zettabytes of data will be generated worldwide (Duarte, 2024).

Table 2. **The amount of data centres by country across the world**.

US	Germany	UK	China	Canada	France	Australia	Netherlands	Russia	Japan
5389	522	517	449	336	314	308	299	251	222

Source: Compiled by the author based on data available at: <https://cloudscene.com/region/datacenters-in-north-america>

In this context, digital data has become an invaluable asset for public and private entities, shaping economic strategies and influencing policy decisions across various sectors.

The growing significance of digital data in shaping economic and policy landscapes underscores the necessity of robust legal frameworks to balance innovation with individual rights, especially concerning the capacity to infer emotions from personal data and utilize such information to influence individual needs (Durovic M., Watson J., *Cambridge Handbook on Artificial Intelligence*, 2022, p. 273). One of the most prominent cases was Cambridge Analytica scandal, where the data of millions of Facebook users was analysed to construct psychological profiles and political orientations, ultimately influencing voter behaviour during the 2016 U.S. presidential election (Hern 2018). With the growing amount of social content generated and capabilities of AI technologies it is difficult to avoid another similar situation.

The privacy risks concerning national security institutions are primarily associated with the mass surveillance capabilities of AI. Under traditional surveillance framework, law enforcement institutions are required to obtain prior authorisation by judicial or administrative authority and provide clear evidence of criminal activities involved. For instance, according to the Criminal Intelligence Law of the Republic of Lithuania, surveillance of correspondence through technical means must be authorized by a court (Article 10). Therefore, the traditional approach to surveillance is characterised by its targeted nature, focusing on specific subjects and ensuring that infringements on privacy are limited by time, location, and scope.

The European Court of Human Rights (further – ECHR) in the case of *Weber and Saravia v. Germany* developed six safeguards, later called the *Weber criteria*, that should be set out in statute law in order to avoid abuses of power:

- the nature of the offences which may give rise to an interception order;
- a definition of the categories of people liable to have their telephones tapped;
- a limit on the duration of telephone tapping;
- the procedure to be followed for examining, using and storing the data obtained;
- the precautions to be taken when communicating the data to other parties;
- the circumstances in which recordings may or must be erased or the tapes destroyed (paragraph 95).

Additionally, in the case *Zakharov v. Russia*, the Court added “reasonable suspicion” through necessity and proportionality test (paragraph 193). These cases prove that to a certain point in time ECHR ruled for strict framework of targeted surveillance (Vardanian and Stehlik, 2022, p. 260).

The COVID-19 pandemic significantly loosened surveillance boundaries, extending monitoring to larger segments of the population. Several surveillance technologies and applications commonly used by national security agencies have been used during the pandemic, including CCTV, facial recognition software, data from mobile phones, financial transactions, and social media intelligence (Davis 2021, 159) This unprecedented use of surveillance tools raised the dilemma within society whether privacy should be sacrificed to achieve a greater sense of security in times of crisis.

Vardanian L. and Stehlik, in their analysis of the ECHR case *Big Brother Watch and Others v. United Kingdom* (2021), observe that the Court notably softened its previously stringent stance on mass surveillance. This shift is marked by the removal of several key parameters for assessing "legality," "necessity in a democratic society," and "proportionality" (p. 261). The Court discarded the principle of ex post facto notification, which would have required informing individuals of their surveillance after its conclusion, thereby denying them the opportunity to exercise effective judicial protection (p. 261). Furthermore, it refused to consider the necessity for prior judicial authorization, emphasizing that a general reference to threats to national security in the applicable legal acts suffices to meet verification requirements (p. 262).

Additionally, the Court imposed no strict requirements for the formulation of legal frameworks governing surveillance, effectively granting states broad discretion in implementing mass surveillance measures (p. 263). Despite criticism, this change in Court’s case law extends the range of mass surveillance under the legal framework of national security suggesting its potential objective as preventive mass surveillance. Given the ongoing Russian aggression in Ukraine and the hybrid warfare targeting nations that support Ukraine, the deployment of such preventive mass surveillance tools appears not only plausible, but potentially already in effect.

The capabilities of AI systems to perform mass surveillance has no doubts and has only technical limitations such as computational power and infrastructure. There is no individual who could surpass an AI system in such data analysis and such capabilities provide national security institutions a tool to monitor public spaces whether physical or digital.

The technical side of the privacy risks is related to data per se. As Peter Norvig, chief scientist at Google admits: “We don’t have better algorithms than anyone else; we just have

more data" (Cleland, 2011). The success of AI systems fundamentally depends on the availability of vast datasets, initially utilized for training algorithms and subsequently employed for content generation or field-specific tasks.

Algorithms themselves do not classify data as personal or non-personal. As a rule, machine learning models do not contain any personal data, only information about groups and classes (Ebers M., *Algorithms and Law*, 2020p. 64). In the case of facial recognition technology, the algorithm is trained to recognize faces, analyse facial features, compare these features against existing databases or online information, and determine potential matches. In this context, privacy concerns primarily would fall on the deployer of the tool, including the legality of the database being accessed, the data processing activities involved, and compliance with laws such as the GDPR or AI Act.

Database creation is a standard practice globally and the national security institutions have all rights to use national information systems and registrars to implement their functions. However, the deployment of facial recognition systems, particularly real-time remote biometric identification systems in publicly accessible spaces, is subject to strict regulation under the AI Act. As a general rule, such systems are classified as prohibited practices (Article 5) and may only be deployed in exceptional cases for law enforcement purposes, primarily to confirm an individual's identity. The criteria, outlines in Article 5 paragraphs 2-7 align with the Weber criteria, established by ECHR and the Courts initial strict approach to privacy protection.

The EU law restricts EU institutions from intervening in the national security affairs of its member states. As a result, privacy risks arising from national security institutions, often primary deployers of prohibited AI practices, may remain undisclosed to society or the individuals affected by such operations, unless a significant information leakage, such as Snowden revelation, would uncover information on privacy violations and would fuel public and political debates on further regulation.

2. Bias and Discrimination

Bias in AI can be defined as algorithmic unfairness, resulting in discrimination. As stated by the UK House of Lords Select Committee on Artificial Intelligence (2018), AI systems are designed to spot patterns during learning process. However, if the data is unrepresentative, or the patterns reflect historical patterns of prejudice, then the decisions made by these systems may also be unrepresentative or discriminatory (Select Committee on Artificial Intelligence, 2017, p. 41).

Daniel Verona and Juan Luis Suarez (2022) in their article outline 6 classifications for bias (p. 3-4):

1. *Sample or selection bias*. Occurs when the **sample** representation is compromised, resulting in significant imbalances;
2. *Measurement bias.* Refers to systematic errors in **data accuracy**, compromising the reliability of values used to support estimations;
3. *Self-reporting (survey) bias*. Relates to **incomplete data** that undermines statistical significance and the accuracy of predictions;
4. *Confirmation (observer) bias*. Arises from a researcher's own prejudices influencing the **presentation of information** to support their working hypothesis;
5. *Prejudice (human) bias*. Occurs when the model or algorithm reflects pre-existing biases inherent in the **knowledge** base used for training;
6. *Algorithm bias*. Occurs when a model or algorithm amplifies bias from the **training dataset** in an effort to address processing demands, often when working with datasets of differing sizes.

These classifications highlight the critical role of data in the design and deployment phases of AI systems. Poor data quality leads to inaccurate results, biases and discrimination in predictions. For instance, predictive policing in the U.S. has led to racial bias, with the number of African Americans detained being disproportionately higher (Heaven, 2021). Regardless of how advanced an AI algorithm may be, it cannot correct inherent issues in flawed data.

EU AI Act stresses non-discrimination importance in all AI systems to be consistent with Charter of Fundamental Rights of the EU and Ethics guidelines for trustworthy AI. AI systems must be developed on the basis of training, validation and testing data sets that meet the quality criteria (Article 10 paragraph 1).

These classifications also underscore the European Union's efforts to establish standards for AI and training data, as articulated in the AI Act. According to Preamble (59) of the AI Act:

"*If an AI system is not trained with high-quality data, does not meet adequate requirements in terms of its performance, accuracy, and robustness, or is not properly designed and tested before being put on the market or otherwise deployed, it may single out individuals in a discriminatory or otherwise incorrect or unjust manner*."

As part of EU's Data Strategy, Regulation (EU) 2023/2854 (Data Act) fosters achievement of high-quality data objective in AI systems and data availability for reuse. This benefits innovation and competitiveness of medium or small enterprises against large stakeholders

or gatekeepers who possess large amounts of data. This Act is oriented to mainly non-private product or related service data (Article 2).

Another significant aspect of non-discrimination is the principle of data minimization, as required under GDPR, which mandates that data collection and processing must be limited to what is necessary for the purpose of the AI system. The principles of data minimisation and data protection by design and by default are essential when processing involves significant risks to the fundamental rights of individuals (Regulation (EU) 2023/2854, Preamble (8)).

However, researchers often collect a wide range of data, much of which may be irrelevant to the predictive outputs. For instance, the example of racial discrimination in predictive policing technologies highlights significant concerns regarding whether race was a critical characteristic influencing a detainee's identification or profiling. Excluding race might have changed patters and outputs.

Therefore, the trustworthiness of AI systems is a key objective embedded in national AI strategies worldwide, aimed at ensuring that such systems are non-discriminatory and substantially tested. Verona D. and Suarez J. L. (2022) suggest that trustworthiness encompasses a range of overlapping properties, including reliability, reproducibility, safety, security, privacy, accuracy, robustness, fairness, accountability, transparency, and explainability (pp. 9–10) This broad definition highlights that the risks associated with AI systems cannot be viewed in isolation, as these risks are interrelated and often overlap.

To enhance trustworthiness in data and promote non-discrimination, nations initially introduced ethics guidelines [6] and are now transitioning toward international standardization. Among the most relevant initiatives addressing bias and discrimination is the ISO/IEC 42001:2023 standard for AI management systems (Grubenmann and Masoni, 2024) and ISO/IEC 23894 or managing risks connected to the development and use of AI (McGarr , 2023). These standards outline requirements for organizations to establish trustworthy AI management practices, including risk management, AI system impact assessment, system lifecycle management, and oversight of third-party suppliers to mitigate potential risks.

Although international standards are non-binding, the stringent requirements of regulatory instruments such as the NIS2 Directive and the AI Act for high-risk AI systems will likely incentivize developers and deployers to adopt both or one of them. Compliance can provide

[6] In EU Ethics Guilines for Trustworthy AI were introduced by High Level Experts Group, In US, Department of Defence adopted ethical principles for AI, in China – Scientific and technological ethics regulation

organizations with international recognition of their trustworthiness and enhance their competitive position in global markets.

Discriminatory decisions are generally attributed to prediction, selection or estimation algorithms (Verona D. and Suarez J. L., 2022, p. 3) which are directly linked to expeditious decision-making processes critical for national security institutions. The effective management of these risks is essential to ensure the quality and integrity of decisions, whether fully automated or merely suggestive in nature. Consequently, national security institutions will have to train their personnel on effective criterion selection and evaluation of potentially discriminatory outputs (predictions) and to adopt ethical principles for trustworthy AI or align their practices with international standards to mitigate risks. Inadequate or false decisions of national security institutions might lead to social unrest and hinder political stability of a country.

3. Infrastructure and Cybersecurity

Global digitalization has transformed every aspect of society, from commerce and communication to governance and infrastructure, thus significantly increasing reliance on network connectivity, making interconnected systems a cornerstone of modern life. When assessing the risks associated with AI applications in infrastructure and cybersecurity, governments must first evaluate their infrastructure's capacity to effectively harness AI technologies and foster innovation. Additionally, they must ensure the cybersecurity of such infrastructure, systems, and the data they store. It is called AI-enabling infrastructure.

However, AI systems requires significant computing power, and generative or general-purpose AI systems might already use around 33 times more energy to complete a task than task-specific software would (Kemene, Valkhof and Tladi, 2024). The establishment of testbeds or sandboxes—controlled environments for evaluating and testing AI technologies—further amplifies energy demands, requiring dedicated energy sources, grids, and related infrastructure.

This challenge is compounded by the competing priorities of sustainable development goals and environmental policies, which emphasize reducing CO2 footprints, while AI leadership objectives drive countries to increase energy supply and consumption.

Sustainable development and environmental policies demand to lessen the carbon footprint where leadership in AI objectives push countries to increase energy supply and consumption. For instance, training GPT-3 (Open AI) is estimated to use just under 1,300 megawatt hours of electricity. This is roughly equivalent to the annual power consumption

of 130 homes in the US. While training the more advanced GPT-4, meanwhile, is estimated to have used 50 times more electricity (Kemene, Valkhof and Tladi, 2024).

US decided to address these challenges through collaborative partnerships with industry, academia, government agencies, and international allies and partners (Executive order 14110, Section 5.2(g)(iv)) aiming to incentivize innovative solutions for opposing objectives. EU's legal acts are of a horizontal nature and it is unclear how resource minimisation (AI Act Article 10) should promote environmental sustainability. As the process of AI embracement just began, there are only preliminary estimates of what might be needed to secure global leadership in transformative technologies.

Another critical aspect of risk associated with infrastructure and cybersecurity is the protection of sensitive information. AI systems, by design, do not inherently classify information as sensitive or non-sensitive. Instead, they analyse large datasets—often incorporating open-source materials—and identify patterns or keywords, grouping data based on algorithmic logic rather than contextual judgment.

The determination of whether specific information is sensitive or requires classification lies solely with human oversight. Without proper human intervention, there is a significant risk that AI systems could inadvertently expose or misuse sensitive information. For example, a little over three dozen security vulnerabilities have been disclosed in various open-source AI and machine learning models, some of which could lead to remote code execution and information theft (Lakshmanam, 2024) which is critical to national security institutions, sensitive information collected and covert operations.

On the other hand, AI systems play a critical role in monitoring digital infrastructure within institutions, identifying discrepancies and vulnerabilities in real-time. However, the implementation of such systems raises significant concerns regarding transparency and accountability, leaving the results to be assessed by human experts.

4. Transparency and accountability

Transparency and accountability risks associated with AI are interdependent variables within the trustworthiness paradigm. According to the transparency obligation, AI systems must ensure traceability, explainability, and effective communication to allow stakeholders to understand how decisions are made and outputs are generated. While, accountability focuses on liability frameworks and requires mechanisms for auditability, minimization and reporting of negative impacts, trade-off management, and redress mechanisms for affected

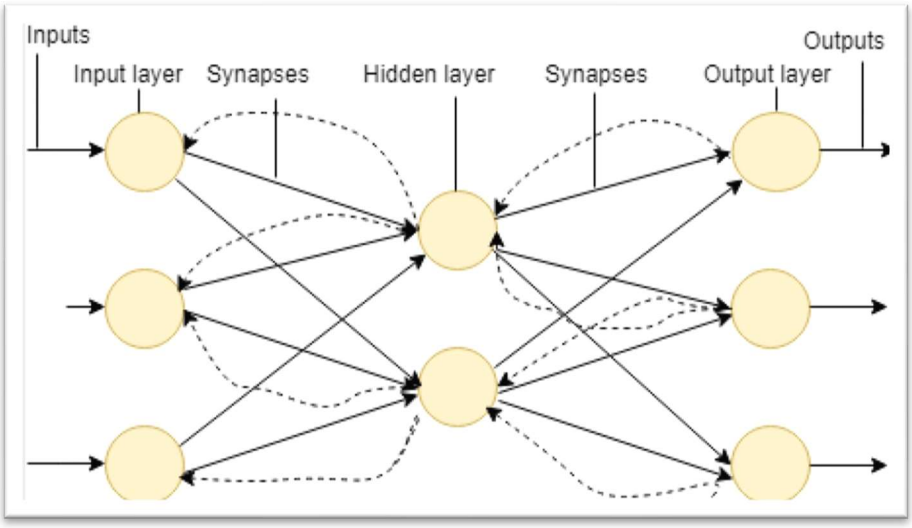
parties (High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019, p. 14).

Transparency concerns focus not only on the data and algorithm, but also on the potential to have some form of explanation for any AI-based determination (*The Cambridge Handbook of Artificial Intelligence*, p. 5). As noted by Pascal D. König et al., opacity—often regarded as the opposite of transparency—can arise either intentionally or from a lack of literacy and expertise in the field (*The Cambridge Handbook of Artificial Intelligence*, p. 30).

According to Martin Ebers (2020), intentional opacity may serve to protect privacy, preserve competitive advantages, safeguard national security interests, or achieve cybersecurity objectives. Conversely, unintended opacity often stems from technical complexity and a lack of institutional expertise (*Algorithms and Law*, p. 49). For instance, while publicly national security institutions may apply intentional opacity for security reasons, internal audits might reveal that these institutions lack the necessary skills to interpret or deploy AI systems effectively. This skills gap is notable due to disparities in salaries and resources between the public and private sectors, particularly in specialized fields such as coding and algorithm analysis.

Another transparency issue relates to the black box effect of algorithms, particularly in artificial neural networks used in deep learning, where only the input and output data is visible, processes occurring within the network remain opaque and difficult to understand (See figure 2). In such a network, all learned information in a neural network is not centralized but is distributed across the network, modifying the architecture the network and the strength of individual connections between neurons (Ebers M., 2020*, Algorithms and Law*, p. 50). This distributed learning mechanism adds an additional layer of complexity and opacity, making it challenging to trace or interpret the decision-making process, necessary for transparency and accountability.

Figure 2. **Black box effect in artificial neural networks.**



Source: State-of-the-art in artificial neural network applications by Abiodun O. I. et al. (2018)

As Pardalos, Rasskazova and Vrahatis (2021) state, there are no algorithms that are optimal for instances across a wide class of problems. Therefore, it is necessary to develop methods that learn from data, identify structures within the objective function, and exploit this knowledge for data-efficient black-box optimization (p. 4).

Currently, the application of these systems is sector-specific, but they have proven to be highly successful in prediction and pattern recognition, which are beneficial to national security institutions. Network data analysis enhances accuracy, processing speed, fault tolerance, latency, performance, volume, and scalability, thereby improving the capabilities of these institutions in handling complex data environments (Pardalos, Rasskazova, Vrahatis, 2021, p. 20).

AI accountability refers to the idea that artificial intelligence should be developed, deployed, and utilized such that responsibility for bad outcomes can be assigned to liable parties (Carnegie Council). While there are ongoing debates regarding the possibility of direct AI liability for wrongdoing, especially as AI systems increasingly outperform human capabilities, current discussions remain grounded in the realities of present AI technology. Given the vulnerabilities of current AI systems and the legal requirements for human-in-the-loop processes, accountability frameworks typically assign liability to developers, providers, or deployers of AI systems. These entities are expected to ensure compliance with applicable regulations.

For instance, the EU's AI Act incorporates accountability through specific provisions related to necessary documentation, process management, and penalties for non-compliance. These provisions are interconnected, forming a framework that ensures responsible development and deployment of AI systems.

For example, the risk management system, as outlined in Article 9, integrates both procedural and documentary obligations. Providers of high-risk AI systems must maintain comprehensive records of risk assessments, mitigation measures, and ongoing monitoring processes. This documentation serves not only as a compliance mechanism but also as an essential tool for demonstrating accountability in the event of adverse outcomes or audits.

Another significant obligation pertains to supply chain security, referred to as the value chain (Article 25). These provisions obligate providers to ensure that all components and services within the supply chain comply with the Act's requirements. This measure complements the NIS2 Directive, collectively enhancing the EU's capacity to safeguard the security and accountability of any element within the supply chain for AI systems.

Transparency and accountability issues, though closely interrelated, may assume different roles within national security institutions. While transparency in the application of AI systems in these institutions is not always a critical component due to the sensitive nature of their operations, accountability requirements cannot be entirely shielded under the guise of national security. The rule of law, enshrined in the constitutions of democratic nations, mandates that all institutions, including those responsible for national security, adhere to applicable legal frameworks. As a result, national security institutions would be required to comply with documentation obligations, facilitating audits or investigations (likely internal) concerning their use of AI systems.

However, in cases where a private individual suffers substantial harm, or the deployment of an AI system results in damage to a group of individuals, these institutions may be compelled to provide the necessary documentation and materials to a competent authority for the adjudication of human rights violations.

5. Economic disruption

Every major economic transition throughout history has been accompanied by widespread societal disruptions, including massive strikes and economic crises, creating tensions between labour forces, businesses, and governments. As Peter P. Groumpos (2021) observes, a revolution is a tumultuous and transformative event, or a series of events and

actions, aimed at fundamentally altering a nation, region, or society, with substantial impacts on the industrial and, more recently, the business world (p. 464).

Until the end of the 20th century, there were four major industrial revolutions: Mechanization, Electrification, Automation, and Digitalization. Among these, Mechanization is considered a pivotal transformation, marking the transition from hand production methods to machine-based production (Groumpos P. P., 2021, p. 465). This period also led to significant socioeconomic reforms, including massive urbanization, the introduction of regulations governing working conditions, and electoral system adjustments in Great Britain, which at the time held a dominant position as the world's leading commercial nation (*BBC bitesize*).

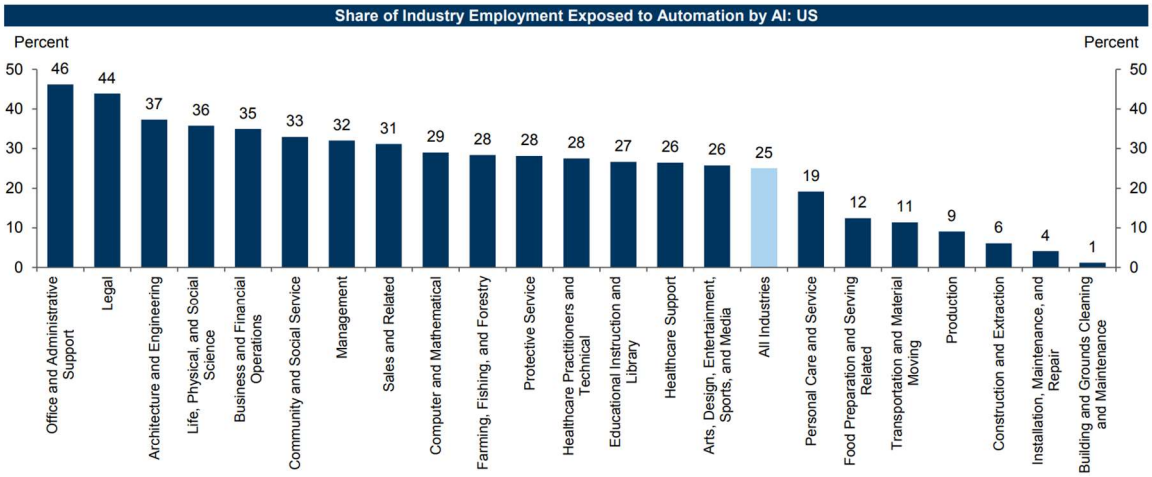
Some experts (CSIS discussions, 2022) characterize AI transformative technologies as the fifth industrial revolution, citing their ubiquitous influence on daily life, their integration across diverse economic sectors, and their recognized potential by leading nations to drive global leadership. However, this transition carries significant socioeconomic risks, including the potential to exacerbate unemployment rates, widen income inequality, and place considerable pressure on social welfare systems.

AI systems, particularly general-purpose models, have already demonstrated capabilities that surpass human performance in big data analysis. While these systems are primarily constrained by computing power, they are increasingly regarded as critical tools for achieving global leadership objectives. However, their transformative potential poses significant economic and regulatory implications, as they have the capacity to automate tasks traditionally performed by humans, potentially displacing both unskilled and skilled labour across a diverse range of industries.

According to Grant Thornton's HR Leaders survey (2024) conducted in the US, 28% of workers reported that their jobs are likely to be reduced or eliminated due to AI adoption, highlighting the potential impact of automation on the workforce. Furthermore, 77% of human resource leaders stated that their organization already has an AI strategy in place, reflecting the rapid incorporation of AI technologies into workplace operations.

US government admits (executive order 14110, Sectio 3(k)) that AI model that is trained on broad data, generally uses self-supervision, contains at least tens of billions of parameters, is applicable across a wide range of contexts substantially lowers the barrier of entry for non-experts. If considering sector specific AI model, such barrier could be even lower.

Figure 3. **Occupation automation rates by AI in US and Europe**



Source: Hatzius et al., 2023, p. 7.

According to Goldman Sachs estimates (2023) approximately 40 percent of all occupations across US and Europe could be subject to automation by AI, with particularly high exposure in administrative roles (46%) and legal professions (44%) (see figure 3).

When considering the necessity of sector-specific expertise, the deployment of AI technologies in national security institutions could enhance the quality of outputs for expeditious decision-making. However, the broader implications of automation and the replacement of occupations could lead to a significant rise in unemployment rates and increase the financial burden on the state. This may compel governments to deplete savings or incur national debt, exacerbating economic vulnerabilities.

Moreover, such economic crises could heighten the operational workload of national security institutions or be exploited by hostile state or non-state actors to disseminate disinformation, destabilize societal cohesion, and incite extremist ideologies.

U.S. policies on artificial intelligence are primarily directed at achieving global leadership and creating opportunities for companies, often leaving individuals to navigate challenges independently, either by adapting to the new landscape or waiting for opportunities to emerge. The European Union, by contrast, adopts a human-centric approach across its policies, emphasizing economic well-being as one of the core principles of the Union.

EU classifies AI systems used in employment, such as those for recruitment, promotion, or termination, as high-risk (Article 6, Annex III) requiring human oversight mechanisms (Article 13) to ensure fairness and accountability in critical employment decisions. However, as the AI Act is novel instrument, its regulatory mechanisms are still formulated to ensure coherence and enforceability. Despite the numerous surveys and challenges raised

in this transitional period, concrete solutions remain limited as the process of adaptation and implementation is still underway.

6. Intentional misuse

The vulnerabilities of AI, which have the potential to result in human rights violations, can be exploited by non-state actors or hostile states. A pertinent example is the ongoing Russian aggression against Ukraine, where hybrid warfare techniques are extensively employed.

Social tensions and the escalating threat of aggression expansion pose significant risks to neighbouring democratic states, rendering societies more vulnerable to propaganda and misinformation. Such societal conditions create opportunities for hostile state or non-state actors to manipulate public opinion and undermine the functionality of national institutions, thereby compromising political governance and social stability.

In May 2024 Open AI company released a report on the use of AI systems for covert influence operations presenting analytical insights into how various actors have utilised their products to support covert influence operations online. They defined covert influence operations as deceptive attempts to manipulate public opinion or influence political outcomes without revealing the true identity or intentions of the actors behind them (p. 3). The report identifies key connections to hostile states, including Russia, China, Iran, and Israel (p. 6), and highlights the following attacker trends:

- Content generation;
- Mixing old and new;
- Faking engagements;
- Productivity gains.

The primary application of OpenAI's products in covert influence operations is content generation. However, the generated content alone does not have the capacity to significantly influence large audiences. For such operations to achieve their intended outcomes, the content must be disseminated through established distribution channels and receive engagement, such as shares and interactions, from human users.

A significant example of AI utilization in hybrid warfare is the online information operation known as Doppelganger, which has been active since February 2022 and is attributed to Russian state actors (EU Disinfo Lab, 2024). This operation targets multiple nations with the objective of undermining international support for Ukraine by demonizing the Ukrainian government through accusations of Nazism and corruption. Additionally, it seeks

to sow divisions within nations supporting Ukraine, propagating narratives that such support is a failing strategy detrimental to civil society.

According to the EU Disinfo Lab analysis (2024), the operation employed tactics combining AI tools with social media dissemination capabilities and visual inputs, such as projecting the blue Star of David onto buildings. These tactics fuelled controversy and confusion, resulting in widespread dissemination across various platforms.

Another hybrid warfare technique is the use of deepfakes, a deep learning technology capable of generating synthetic media or content designed to mislead and manipulate public perception. Such synthetic content can distort societal understanding and exploit emotional reactions. Even if the synthetic nature of the content is later debunked, the emotional impact often lingers, potentially influencing public opinion or sparking unrest in unrelated contexts. A significant example is a fake and heavily manipulated video depicting Ukrainian President Volodymyr Zelenskyy, telling his soldiers to lay down their arms and surrender the fight against Russia (Allyn, 2022). It constituted a hybrid operation against Ukraine, as the synthetic content was disseminated not only through social media platforms but was also broadcasted on Ukraine 24 television after the channel's systems were compromised by hackers.

US also recognizes the capabilities of dual-use foundational models to significantly lower the barrier of entry for non-experts in designing, synthesizing, acquiring, or utilizing chemical, biological, radiological, or nuclear (CBRN) weapons, categorizing this as a national security risk (Executive Order 14110, Section 3(k)(i)). As previously noted, AI systems possess the capacity to replace certain occupations due to their advanced technological, data-processing, and analytical capabilities. And these capabilities may be exploited in both ways.

One more example of AI misuse is the intentional exploitation of AI systems by creative individuals. Unlike hostile state actors, these individuals do not seek to disrupt political systems or incite social unrest. Instead, they are often motivated by a desire to test the boundaries of AI capabilities, sometimes leading to unintended consequences or ethical dilemmas. Such creative infringements can result in financial losses for companies or institutions that deploy AI systems. For instance, DPD disabled a portion of its online support chatbot after it used profanity toward a customer and made critical remarks about the company.

National security institutions must remain vigilant regarding the misuse of AI systems, whether intentional or creative. Intentional misuse, such as the deployment of deepfakes, can significantly manipulate public opinion or incite social unrest, depending on the nature

and dissemination of the synthetic content. Institutions must possess the technological capabilities to respond expeditiously to such threats in order to safeguard the well-being of society. Furthermore, technological literacy should be enhanced through governmental and public initiatives aimed at promoting critical thinking over emotional reactions, thereby increasing societal resilience to misinformation and manipulation.

7. AI application risks for national security

All AI-related risks present challenges for national security institutions, both internally and externally, necessitating the development and implementation of comprehensive strategies to mitigate their potential impact. While the prohibition of AI applications may safeguard sensitive information and provide a short-term perception of security, such measures are not sustainable in the long term. As with previous industrial revolutions, the transformative nature of AI will inevitably influence institutional capabilities and compel national security institutions to integrate AI systems into their operational frameworks and daily functions.

Today, one of the most challenging risks is deepfake technology, which is advancing alongside other emerging technologies. In the near future, distinguishing between authentic digital content and synthetic material may become increasingly difficult. This technology possesses wide-ranging application capabilities, spanning from creative uses to manipulative purposes, including the potential to induce social crises or disrupt public trust for the benefit of hostile states during periods of conflict or geopolitical tension.

Hybrid warfare techniques employed by Russia against states supporting Ukraine have also exposed cybersecurity vulnerabilities in both the public and private sectors (See Annex I). This underscores the significance of the EU's NIS2, which is well-timed to address these challenges. The directive aims to enhance overall cybersecurity capabilities, strengthen the resilience of critical infrastructure, and encourage the development and deployment of cybersecurity-oriented AI systems.

Considering privacy risks, the balance between the right to privacy and the protection of societal interests will likely tilt in favour of society, as the interests of the collective are often deemed to outweigh those of the individual. Consequently, the majority of legal instruments incorporate national security exceptions to address such scenarios. However, even within the framework of national security, protective surveillance employing AI systems must adhere to strict legal standards and undergo proportionality and necessity assessments to uphold the rule of law and ensure that such measures are neither arbitrarily nor unjustifiably applied.

To sum up, legislators in democratic states must prioritize the assessment of potential risks over the immediate opportunities presented by AI systems to ensure the preservation of human-centric values. However, it is equally critical not to deprive national security institutions of the potential of AI to safeguard these values. Failure to do so could allow hostile state actors to achieve superiority in informational operations, thereby undermining societal cohesion and destabilizing political systems.

CONCLUSIONS

1. AI technologies, with their transformative capabilities, are globally recognized as having a profound impact. Consequently, every leading economy seeks to establish a regulatory framework to govern AI's development and manage potential negative outcomes effectively.
2. The analysis of global regulatory trends among leading economies reveals distinct approaches to AI governance shaped by their respective political contexts. The EU, as a supranational organization, prioritizes balancing human-centric values with fostering innovation. The US, on the other hand, focuses on addressing AI-related challenges to national security and maintaining its global technological leadership. China, operating within a non-democratic political framework, already wields significant influence in global digital infrastructure, including next-generation cellular networks, fibre optic cables, undersea cables, and data centres, positioning itself as a peer competitor to the US in this transformative domain.
3. While the GDPR has achieved global recognition for setting a high standard in personal data protection, the AI Act's risk-based approach is already being criticized as a potential impediment to innovation. Coupled with the EU's extensive regulatory obligations and lengthy legislative procedures, this framework may hinder the Union's ability to secure a leading position in technological transformation in the near future. Nonetheless, the value-based approach may yield significant results over the long term, providing stability, clarity, and a solid foundation for trustworthy and ethical AI development within the region.
4. Technological capabilities offered by AI are increasingly regarded as superior to traditional intelligence-gathering methods in national security operations, providing enhanced speed, efficiency, and informational superiority. Given the fact that hostile states or non-state actors will exploit such capabilities for their own strategic advantages, EU legislators cannot afford to deprive national security institutions of these transformative tools, despite the inherent risks posed by AI systems.
5. Critical situations such as the 2015 migration crisis or Russia's unprovoked aggression against Ukraine have highlighted a trend where Member States voluntarily relinquish their sovereign right to address internal challenges independently. Once such competences are regulated at the EU level, they must align with EU regulatory provisions, thereby contributing to the establishment of a supranational security framework. This framework allows the EU to intervene

directly in areas traditionally reserved for national security. Consequently, despite the national security exclusions embedded in EU regulatory instruments, these provisions are indirectly applied and must be adopted by national security institutions across the EU.

6. The most prominent risks associated with national security operations, informational superiority, and expeditious decision-making include cybersecurity vulnerabilities, infrastructure risks, intentional misuse of AI technologies, and, to a certain extent, issues of bias and discrimination. To effectively address these risks, national security institutions must prioritize investments in robust infrastructure, field-specific AI technologies tailored to operational needs, and the development of expertise.
7. Privacy, transparency, and accountability risks are critical considerations in the private sector, where they significantly influence consumer trust and regulatory compliance. However, in national security operations, these risks are subject to proportionality assessments, ensuring that any measures implemented are necessary, appropriate, and justified under the rule of law. In such contexts, the principle of collective security is typically prioritized over individual rights, particularly when addressing pressing threats to public safety or national stability.

LIST OF REFERENCES

Legal acts:

1. Charter of the United Nations (1945) International Law Documents: 15th edition. Blackstone's Statutes: Oxford University Press, 2021;
2. Convention on the Rights and Duties of States (1933), International Law Documents: 15th edition. Blackstone's Statutes: Oxford University Press, 2021;
3. Consolidated version of the Treaty on the Functioning of the European Union - PROTOCOLS - Protocol (No 2) on the application of the principles of subsidiarity and proportionality, OJ C 115, 9.5.2008, p. 206–209;
4. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689;
5. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), OJ L, 2023/2854;
6. Regulation (EU) 2021/887 of the European Parliament and of the Council of 20 May 2021 establishing the European Cybersecurity Industrial, Technology and Research Competence Centre and the Network of National Coordination Centres, OJ L 202, 8.6.2021, p. 1–31;
7. Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151, 7.6.2019, p. 15–69;
8. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, p. 1–88;
9. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and

repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJ L 333, 27.12.2022, p. 80–152;

10. Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119, 4.5.2016, p. 89–131;
11. Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, OJ L 194, 19.7.2016, p. 1–30;
12. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, No. COM/2021/118 on 2030 Digital Compass: the European way for the Digital Decade;
13. Communication from the Commission on 25 August 2018 to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, No. COM/2018/237 on Artificial Intelligence for Europe;
14. Executive Order 14110 of the President of the United States of America of 30 October 2023. Federal Register, Vol. 88, No, 210;
15. National Security Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfil National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence, DCPD-202400945
16. New Generation Artificial Intelligence Development Plan (2017) [online] Available at: https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm [Accessed 22 November 2024];

Special literature:

17. Abiodum O., I. (2018) State-of-the-art in artificial neural network applications: A survey. *Heliyon* 4 e00938. DOI: 10.1016/j.heliyon.2018. e00938;
18. Algorithms and Law (2020) ed. by Ebbers M. and Navas S., Cambridge University Press;

19. Barak A. Proportionality: constitutional rights and their limitations, Cambridge University Press, 2012;
20. Buckland B. S., Schreier F., Wincler T. H. (2015) Democratic Governance Challenges of Cyber Security. *DCAF Horizon 2015 Working Paper*, No. 1 [online] Available at: Democratic Governance Challenges of Cyber Security [Accessed 29 November 2024];
21. Cambridge Handbook of Artificial Intelligence (2022) ed. by DiMatteo L. A., Poncibo C., Cannarsa M. Cambridge University Press, DOI: 10.1017/9781009072168;
22. Casolari F. (2023) Supranational Security and National Security in the Light of the EU Strategic Autonomy Doctrine: The EU-Member States security Nexus Revisited, *European Foreign Affairs Review* vol. 28, no. 4, pp. 323-340;
23. Congyan C. (2017) Enforcing a New National Security? China‘s National Security Law and International Law. *Journal of East Asia and International Law*. 10. 4-4. DOI: 10.14330/jeail.2017.10.1.04;
24. Crafts N. (2021) Artificial Intelligence as a general-purpose technology: an historic perspective. *Oxford Review of Economic Policy*, 37(3), Pages 521–536, https://doi.org/10.1093/oxrep/grab012;
25. Davis J. (2021) Surveillance, intelligence and ethics in a COVID-19 world. Section in a book *National Security Intelligence and Ethics*, ed. By Miller S., Regan M., Walsh P. F., p. 156–166, DOI: 10.4324/9781003164197-13;
26. European Council (2024) Strategic Agenda 2024-2029 [online] Available at: sn02167en24_web.pdf [Accessed 23 November 2024];
27. Fein R. A., Vossekuil B. (2000) Protective Intelligence Treat Assessment Investigations. Research report, U. S. Department of Justice;
28. Goldfarb B. (2011) General and Miscellaneous: Book review of Economic Transformations: General Purpose Technologies and Long-Term Economic Growth by Lipsey et all. *The Journal of Economy History*, Vol 71(3) pp. 820-823;
29. Groumpos P., P. (2021) A Critical Historical and Scientific Overview of all Industrial Revolutions. IFAC-PapersOnLine, vol 54(13) p. 464-471, https://doi.org/10.1016/j.ifacol.2021.10.492;
30. Kalodanis K., Rizomiliotis P., Anagnostopoulos D. (2023) European Artificial Intelligence Act: an AI security approach. *Information and Computer Security*, vol 32, No. 3, DOI: 10.1108/ICS-10-2022-0165;

31. Lippman W. (1943) U. S. Foreign Policy: Shiled of the Republic. USA: The Atlantic Monthly Press;
32. Liu Y. 2024, Generative AI: Catalyst for Growth or Harbinger of Premature De-Professionalisation? Policy Research Working Paper for World Bank Group. [online] Available at: https://documents1.worldbank.org/curated/en/099520009172451039/pdf/IDU1aa745fd01bcf014ac51b11d1e9f762ce51e5.pdf [Accessed 16 November 2024];
33. Montasari, R. (2023). Internet of Things and Artificial Intelligence in National Security: Applications and Issues. In: Countering Cyberterrorism. Advances in Information Security, vol 101. Springer, Cham. https://doi.org/10.1007/978-3-031-21920-7_3;
34. National Treat Assessment of the Republic of Lithuania (2024) [online] Available at: GR-2024-02-15-EN-1.pdf [Accessed 22 November 2024];
35. Pardalos P., Rasskazova V., Vrahatis M., N. (2021) Black Box Optimization, Machine Learning, and No-Free-Lunch Theorems, DOI:10.1007/978-3-030-66515-9;
36. Perry W., Signori D., Boon J. (2004) Exploring Information Superiority. National Research Defence Institute, Available at: Exploring Information Superiority: A Methodology for Measuring the Quality of Information and Its Impact on Shared Awareness [Accessed 1 December 2024];
37. Select Committee on Artificial Intelligence (2017) AI in the UK: ready, willing and able? Report of session 2017-19, [online], Available at: AI in the UK: ready, willing and able [Accessed 14 December 2024];
38. Sivan-Sevilla I. (2023), Supranational Security states for national security problems: governing rules and capacities in tech-driven security spaces, *Journal of European Public Policy*, 30:7, 1353-1378, DOI: 10.1080/13501763.2023.2172063;
39. Turing A., M. (1950) Computing Machinery and Intelligence. *Mind 49*, 433-460. Available at: https://courses.cs.umbc.edu/471/papers/turing.pdf [Accessed 16 November 2024];
40. U.S. Department of Defence (2023) Military and Security Developments Involving the People's Republic of China. Annual Report to Congress [online] Available at: 2023 Report on the Military and Security Developments Involving the People's Republic of China (CMPR) [Accessed 22 November 2024];

41. Vardanyan L., Stehlík V. (2022) Is the Case Law of ECtHR Ready to Prevent the Expansion of Mass Surveillance in the Post-Covid Europe? *European Studies* – Vol. 7/2020, DOI: 10.2478/eustu-2022-0056;
42. Varona, D., & Suárez, J. L. (2022). Discrimination, Bias, Fairness, and Trustworthy AI. *Applied Sciences*, 12(12), 5826. https://doi.org/10.3390/app12125826;
43. Wolfers A., (1962) Discord and Collaborations: Essays on International Politics. Baltimore, The Johns Hopkins Press, Available at: Discord And Collaboration Essays On International Politics : Arnold Wolfers : Free Download, Borrow, and Streaming : Internet Archive [Accessed 1 December 2024];

Case law:

44. *Weber and Saravia v. Germany* [ECHR], No. 54934/00 [2000-01-10]. ECLI:CE:ECHR:2006:0629DEC005493400;
45. *Zakharov v. Russia* [ECHR], No. 47143/06 [2006-10-20]. ECLI:CE:ECHR:2015:1204JUD004714306;
46. Privacy international (BCD case) [CJEU], No. C-623/17, [2020-10-06]. ECLI:EU:C:2020:790;

Other sources:

47. Allyn B. (2022) Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn [online] (modified 2022-03-16), Available at: A deepfake video showing Volodymyr Zelenskyy surrendering worries experts : NPR [Accessed 12 December 2024];
48. Barak Obama interview (2016) What AI means for national security [online] (modified 2016-10-12), Available at: President Barack Obama on What AI Means for National Security | WIRED [Accessed 14 November 2024];
49. BBC bitesize, Social and political reforms during the Industrial Revolution [online] (modification date not indicated), Available at: Social and political reform - BBC Bitesize [Accessed 13 December 2024];
50. Britannica. History of Artificial Intelligence [online] (modified 2024-12-13), Available at: History of artificial intelligence | Dates, Advances, Alan Turing, ELIZA, & Facts | Britannica [Accessed 14 December 2024];
51. Carnegie Council, AI Accountability [online] (modification date not indicated), Available at: AI accountability | Carnegie Council for Ethics in International Affairs [Accessed 11 December 2024];

52. Cleland S. (2011) Google's "Infringenovation" Secrets, Article in Forbes [online] (modified 2011-10-04), Available at: Google's "Infringenovation" Secrets [Accessed 7 December 2024];
53. Cloudscene (2024) Amount of data centres worldwide [online] (modified 2024), Available at: North America | Data Center Market Overview | Cloudscene [Accessed 7 December 2024];
54. Coursera. The History of AI: A Timeline of Artificial Intelligence [online] (modified 2024-10-25), Available at: The History of AI: A Timeline of Artificial Intelligence | Coursera [Accessed 14 November 2024];
55. CSIS discussions (2022) National Security and Artificial Intelligence: Global Trends and Challenges [online] (modified 2022-06-28), Available at: National Security and Artificial Intelligence: Global Trends and Challenges [Accessed 13 December 2024];
56. Cyber Peace Institute (2022) Case study: Viasat [online] (modified 2022-06), Available at: Case Study: Viasat Attack | CyberPeace Institute [Accessed 1 December 2024];
57. Dartmouth Summer Research Project: The Birth of Artificial Intelligence [online] (modified 2021-09-30), Available at: Dartmouth Summer Research Project: The Birth of Artificial Intelligence - History of Data Science [Accessed 14 November 2024];
58. Duarte F. (2024) Amount of Data Created Daily [online] (modified 2024-06-13), Available at: Amount of Data Created Daily (2024) [Accessed 7 December 2024];
59. Environment policy: general principles and basic framework (2024) Fact Sheets on the European Union [online] (modified 2024-04), Available at: Environment policy: general principles and basic framework | Fact Sheets on the European Union | European Parliament [Accessed 1 December 2024];
60. EU AI Act: EU AI Act: How risk is classified (2024), Article on Trail [online] (modified 2024-07-30), Available at: EU AI Act: Risk-Classifications of the AI Regulation [Accessed 14 December 2024];
61. EU Disinfo Lab (2024) What is the Doppelganger operation? List of resources [online] (modified 2024-10-30), Available at: Threat Intel Report [Accessed 12 December 2024];
62. European AI Office (2024) [online] (modified 2024-12-12), Available at: European AI Office | Shaping Europe's digital future [Accessed 14 December 2024];

63. European Commission (2018) A definition of Artificial Intelligence: main capabilities and scientific disciplines [online] (modified 2024-12-13), Available at: A definition of Artificial Intelligence: main capabilities and scientific disciplines | Shaping Europe's digital future [Accessed 17 November 2024];
64. European Commission. European Approach to artificial intelligence [online] (Modification date not indicated), Available at: European approach to artificial intelligence | Shaping Europe's digital future [Accessed 14 November 2024];
65. European Commission (2020) Shaping Europe's Digital Future [online] (Modification date not indicated), Available at: 84c05739-547a-4b86-9564-76e834dc7a49_en [Accessed 19 November 2024];
66. European Commission. European Approach to artificial intelligence [online] (Modification date not indicated), Available at: European approach to artificial intelligence | Shaping Europe's digital future [Accessed 14 November 2024];
67. European Commission (2020) Shaping Europe's Digital Future [online] (Modification date not indicated), Available at: 84c05739-547a-4b86-9564-76e834dc7a49_en [Accessed 19 November 2024];
68. Gerken T. (2024) DPD error caused chatbot to swear at customer [online] (modified 2024-01-19), Available at: DPD error caused chatbot to swear at customer [Accessed 12 December 2024];
69. Grant Thornton (2024) HR leaders double down on attraction and retention priorities: survey [online] (modified 2024-08-01), Available at: Grant Thornton survey: HR leaders double down on attraction and retention priorities | Grant Thornton [Accessed 13 December 2024];
70. Grubenmann R. P., Masoni F. (2024) ISO/IEC 42001: The latest AI management system standard [online] (modified 2024-10-04), Available at: ISO/IEC 42001: The latest AI management system standard [Accessed 9 December 2024];
71. Hatzius J. et al. (2023) The Potentially Large Effects of Artificial Intelligence on Economic Growth. Goldman Sachs Economics Research, [online] (modified 2023-03-26), Available at: Global Economics Analyst The Potentially Large Effects of Artificial Intelligence on Economic Growth (BriggsKodnani) [Accessed 13 December 2024];
72. Heaven W. D. (2021) Predictive policing is still racist—whatever data it uses. Article in MIT Technology Review [online] (modified 2021-02-05), Available at: Training data that is meant to make predictive policing less biased is still racist | MIT Technology Review [Accessed 9 December 2024];

73. Hern A. (2018) Cambridge Analytica: how did it turn clicks into votes? Article on The Guardian [online] (modified 2018-05-06), Available at: Cambridge Analytica: how did it turn clicks into votes? | Big data | The Guardian [Accessed 7 December 2024];
74. IBM. The Games that Helped AI Evolve [online] (modification date not included), Available at: https://www.ibm.com/history/early-games [Accessed 15 November 2024];
75. Interesse G. (2024) China Releases New Draft Regulations on Generative AI [online] (modified 2024-05-30), Available at: China Releases New Draft Regulations for Generative AI [Accessed 22 November 2024];
76. Janonis T. (2024) Oficialus atsakymas: ir toliau kategoriškai atsisakoma suteikti informaciją, ar Seime lankėsi Germanas ir Trinkūnaitė. Article in Delfi [online] (modified 2024-11-08), Available at: Oficialus atsakymas: ir toliau kategoriškai atsisakoma suteikti informaciją, ar Seime lankėsi Germanas ir Trinkūnaitė - Delfi [Accessed 25 November 2024];
77. Kemene E., Valkhof B., Tladi T. (2024) AI and energy: Will AI help reduce emissions or increase demand? Here's what to know. Article in World Economic Forum website [online] (modified 2024-07-22), Available at: AI and energy: Will AI reduce emissions or increase demand? | World Economic Forum [Accessed 11 December 2024];
78. Lakshmanam R. (2024) Researchers Uncover Vulnerabilities in Open-Source AI and ML Models [online] (modified 2024-09-29), Available at: Researchers Uncover Vulnerabilities in Open-Source AI and ML Models [Accessed 11 December 2024];
79. McGarr T. (2023) ISO/IEC 23894 – A new standard for risk management of AI [online] (modified 2023-02-24), Available at: ISO/IEC 23894 – A new standard for risk management of AI - AI Standards Hub [Accessed 9 December 2024];
80. McGerty F. (2024) European defence spending: a decade of growth [online] (modified 2024-11-07), Available at: European defence spending: a decade of growth [Accessed 1 December 2024];
81. Milmo D (2024) Meta pulls plug on release of advanced AI model in EU. Article on The Guardian [online] (modified 2024-07-18), Available at: Meta pulls plug on release of advanced AI model in EU | Meta | The Guardian [Accessed 15 November 2024];

82. Ministry Of National Defence of the Republic of Lithuania. Cybersecurity [online] (modified 2024-01-11), Available at: LR Krašto apsaugos ministerija [Accessed 29 November 2024];
83. Open AI (2024) AI and Covert Influence Operations: Latest Trends [online] (modified 2024-05), Available at: Threat Intel Report [Accessed 12 December 2024];
84. The Brussel Times (2024) ChatGPT consumes 25 times more energy than Google [online] (modification date not indicated), Available at: ChatGPT consumes 25 times more energy than Google [Accessed 22 November 2024],

Additional information on AI use:

ChatGPT was used to check legal vocabulary and certain ideas, in order to have a better outlook.

SUMARY

Legal Aspects of Application of Artificial Intelligence in National Security

Lina Sokol

Advancements in artificial intelligence (AI) are widely regarded as one of the key pillars of global leadership within the international community. As a foundational technology, AI possesses the capacity to drive transformative innovation and accelerate industrial growth. Consequently, legislators in leading global economies recognise the necessity of implementing robust regulatory frameworks to ensure effective AI governance while balancing innovation and societal impact.

Global regulatory trends provide different approaches towards AI governance. US concerns are related to AI challenges for national security and preservation of global leadership, In China AI governance is fully regulated by the ruling party and national security institutions. China, already possess significant capabilities in the global digital infrastructure and data collection. EU takes a human-centric approach which is visible through all related regulatory instruments – GDPR, NIS2 and AI Act.

National security matters are exclusive competence of EU member states and states have freedom of choice to reach informational superiority and provide political leaders with timely qualitative outputs. Capabilities of AI technologies allow national security institutions to achieve their goals expeditiously. However, in critical situations member states voluntarily give up their sovereign right to address internal problems individually. Once regulated on the EU level they become part of the EU legislature and supranational instrument.

AI technologies, though in early stage of transformation, pose significant risks. However, only some of them are relevant to national security institutions – infrastructure and cybersecurity, intentional misuse and to certain extent – bias and discrimination. Other risks are more relevant to private sector to influence consumer trust and regulatory compliance.

ANNEX I

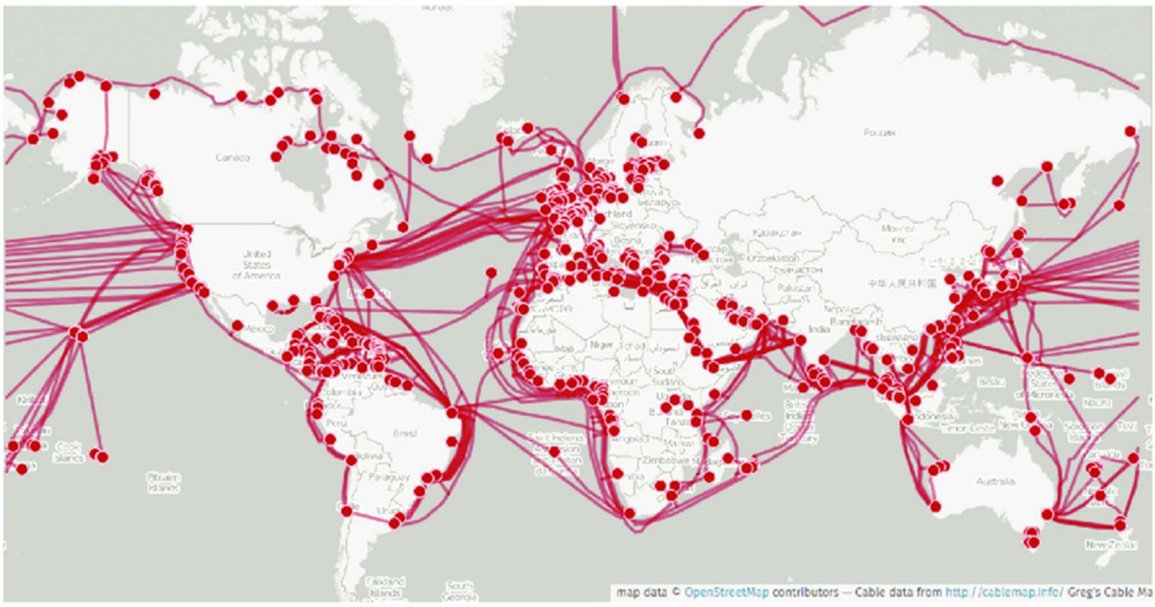
List of notable cyber incidents across the world (listed and organised by Chat-GPT from OpenAI)

Year	Incident	Target	Impact	Significance
2010	Stuxnet	Iran's nuclear enrichment facilities	Damaged centrifuges, delayed Iran's nuclear program	First cyber weapon to cause physical damage, showcasing potential of cyber warfare
2011	Sony PlayStation Network Hack	Sony PlayStation Network	Compromised personal info of 77M users; 23-day service outage	One of the largest breaches in gaming history
2013	Target Data Breach	Target Corporation	Stolen credit/debit card info of 40M customers; personal info of 70M more	Raised awareness of retail and payment system security
2013–14	Yahoo Data Breaches	Yahoo user accounts	Compromised 3 billion user accounts	Largest data breach in history; impacted Yahoo's reputation and value
2014	Sony Pictures Hack	Sony Pictures Entertainment	Exposed internal emails, unreleased films, and employee data; disrupted operations	Allegedly orchestrated by North Korea in retaliation for *The Interview*, highlighting geopolitical motives in cyberattacks
2015	OPM Breach	U.S. Office of Personnel Management	Stole data of 21M U.S. government employees and applicants	Major espionage operation, allegedly by Chinese hackers
2015–16	Ukraine Power Grid Attack	Ukrainian power grid	Power outages affecting hundreds of thousands	First known cyber attack on a power grid; attributed to Russian actors
2017	WannaCry Ransomware	Global	Locked files on 200,000+ computers in 150 countries; affected UK's NHS	Exploited Microsoft vulnerability; underscored need for software updates and ransomware defenses
2017	NotPetya	Primarily Ukraine; spread globally	Caused billions in damages; disrupted operations of Maersk, Merck, FedEx	State-sponsored (attributed to Russia); used as geopolitical tool
2017	Equifax Data Breach	Equifax	Exposed personal and financial data of 147 million people	One of the largest breaches of sensitive consumer information

Year	Incident	Target	Impact	Significance
2018	Marriott Data Breach	Marriott International	Exposed personal information of 500M guests	Highlighted vulnerabilities in hospitality data systems
2018	Cambridge Analytica Scandal	Data of Facebook users	Misused data of 87M users to influence political campaigns	Raised ethical concerns around data privacy and social media's role in democracy
2019	Baltimore Ransomware Attack	City of Baltimore, Maryland, USA	Disrupted city services; ransom demanded	Large ransomware attack on a U.S. city; underscored vulnerabilities of municipal systems
2020	SolarWinds Attack	SolarWinds software	Compromised U.S. federal agencies and private sector companies	Sophisticated supply chain attack; exposed vulnerabilities in third-party software
2021	Colonial Pipeline Attack	Colonial Pipeline	Caused fuel shortages and panic buying on U.S. East Coast	Demonstrated ransomware's impact on critical infrastructure; spurred cybersecurity initiatives in energy sector

ANNEX

World Undersea Cable Map by Dwayne Woods and Junda Li[7]



7 https://www.researchgate.net/publication/385906808_Asian_Security_ISSN_Print_Online_Journal_homepage_wwwtandfonlinecomjournalsfasi20_Dangerous_depths_of_bifurcation_the_rise_of_inter-_national_security_narcissists_and_undersea_cable_dis_connections_Da (2024-12-07)