UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

The Department of Political Science, Law and International Studies

Master's degree program in European and global studies LM-90, Global social policies and security issues track

Final thesis

The Geopolitical Impact of the AI Act on the Global Sphere: A Comparison the EU Regulation of Artificial Intelligence with the experience of China and Russia

Instructor:
Prof. Daniele Ruggiu

Student:
Aleksandra Nikitina
Student number: 2071480

Academic year 2023/2024

Table of contents:

Abstract:

The European Union Artificial Intelligence Act (EU AI Act) stands as a pivotal milestone in governance of AI at the global level, underscoring the EU's ascendancy as a regulatory force in the digital realm. This work aims at offering a comprehensive examination of the EU AI Act, with a focus on its historical genesis, regulatory architecture, and geopolitical ramifications. Furthermore, this thesis will undertake a comparative analysis, comparing the regulatory landscape of the EU AI Act with the authoritarian experiences of China and Russia in the field of AI governance. Through this comparative lens, potential threats to individual liberties and societal well-being will be highlighted, offering valuable insights into the divergent approaches to AI regulation. Lastly, this work will culminate in an evaluation of the EU AI Act, contextualizing its significance within the broader geopolitical landscape. By assessing the Act's implications for geopolitical power dynamics and international relations, this work seeks to provide a nuanced understanding of its geopolitical ramifications.

Introduction:

The advent of artificial intelligence (AI) has revolutionized diverse sectors, prompting global powers to establish comprehensive regulatory frameworks to harness its potential while mitigating risks. The European Union Artificial Intelligence Act (EU AI Act) stands as a pivotal milestone in AI governance, underscoring the EU's role as a leading regulatory force in the digital realm. This thesis aims to offer a comprehensive examination of the EU AI Act, focusing on its historical genesis, regulatory architecture, and geopolitical ramifications. Additionally, a comparative analysis will be conducted, contrasting the EU AI Act with the AI governance frameworks of authoritarian regimes such as China and Russia. Through this comparative lens, potential threats to individual liberties and societal well-being will be highlighted, providing valuable insights into divergent approaches to AI regulation.

The research is anchored on the following question: "How do the regulatory frameworks for artificial intelligence in the European Union, Russia, and China differ, and what are the implications of these differences for societal impacts and geopolitical dynamics?" This research employs a qualitative comparative analysis to examine the societal impacts of AI regulations within democratic and authoritarian regimes, focusing on the European Union (EU), Russia, and China. The selected cases represent contrasting governance structures: the EU, with its democratic framework and comprehensive AI regulation aimed at balancing innovation and ethical standards, and Russia and China, where AI is leveraged for control and surveillance, reflecting different regulatory priorities and societal implications. The analysis covers regulatory developments and policies from 2018 to 2024, with a geographical focus on the EU, Russia, and China. This timeframe captures recent advancements and implementations of AI regulations, providing a current perspective on their impacts. The qualitative approach allows for an in-depth understanding of regulatory texts, policy documents, and their contextual applications. By employing an inductive method, the research begins with the analysis of regulatory frameworks in the selected regions and connects these observations to broader theories on governance, technology, and societal change. Primary sources include AI policy documents, regulatory texts, and official guidelines from the EU, Russia, and China, supplemented by secondary literature such as scholarly articles and policy analyses. The research involves document collection and content analysis to identify key themes, principles, and objectives of the AI regulations. Additionally, interviews with legal experts from Russia and China ensure accurate interpretation and

understanding of the respective AI legislation, providing practical insights into their application and implications. These interviews, mainly conducted in the second and third chapters, follow a semi-structured format to allow for in-depth exploration of complex issues. Given the sensitive nature of these interviews, particularly involving individuals who might be considered political opponents because of the provided information, several measures were adopted to protect personal data and ensure confidentiality:

1. **Informed Consent**: Each interviewee was provided with an informed consent form outlining the purpose of the research, the use of their data, and their rights to withdraw at any time.
2. **Data Protection**: Personal data, including the informed consent forms, were stored in a separate file protected with a password in order to protect sensitive data of respondents. Additionally, their names were purposely changed to ensure safety.
3. **Physical Security**: If data was stored on a portable device such as a pen drive, the file containing the work was protected with a password to prevent unauthorized access in case of loss.
4. **Digital Security**: Research work was conducted in secure environments, avoiding the use of applications that collect data to minimize data vulnerabilities.

These steps are crucial to ensure the ethical handling of sensitive data and to maintain the integrity of the research process.

The comparative analysis framework evaluates the AI regulations based on criteria such as ethical considerations, control mechanisms, innovation support, and societal impact. By synthesizing the findings, the research draws conclusions about the broader impact of AI regulation in democratic and authoritarian contexts, identifying potential threats and opportunities within the global geopolitical landscape. The significance of this analysis lies in its comprehensive comparative approach, which illuminates the distinct ways in which democratic and authoritarian regimes regulate AI and the resulting societal impacts. By examining the European Union (EU), Russia, and China, the study provides valuable insights into how different governance structures shape the deployment and oversight of AI technologies. The findings aimed at understanding the ethical, social, and political dimensions of the current AI regulation, highlighting the EU's efforts to balance innovation with ethical standards, contrasted with Russia's and China's use of AI for control and surveillance.

However, the analysis also has its limitations. Firstly, the study is inherently constrained by the availability and quality of publicly accessible regulatory documents and policy analyses, which may not fully capture the intricacies and enforcement nuances of AI regulations in each region. Secondly, while interviews with legal experts from Russia and China provide valuable insights, they are limited in number and scope, potentially omitting diverse perspectives within those countries. Additionally, the research focuses on a specific timeframe (2018-2024), which, although capturing recent developments, may miss long-term trends and the evolution of AI policies beyond this period. Furthermore, the qualitative nature of the study, while providing depth, may lack the quantifiable metrics that a more data-driven approach could offer. These limitations suggest that the findings should be interpreted as part of a broader, ongoing discourse on AI regulation, necessitating further research and continuous monitoring to fully understand the dynamic and multifaceted impacts of AI on society.

The strategic positioning and regulatory approaches to artificial intelligence (AI) have become focal points for global competitiveness. This literature review explores the divergent AI strategies and governance frameworks adopted by various regions, emphasizing the distinct cultural, political, and economic contexts that shape these approaches. Key areas of comparison include the ethical principles, regulatory philosophies, and strategies of the integration of AI within national development goals.

Following the analysis of Brattberg, Rugova, and Csernatoni (2020), the findings of the thesis reveal despite its strong industrial base and leading AI research, the EU underperforms compared to the United States and China due to market fragmentation and challenges in attracting investment. Therefore, considering its limited political weight, the EU aims to leverage its regulatory power to promote "trustworthy AI," balancing innovation with ethical human-centric considerations (Brattberg et al., 2020, p. 1) by building consumer confidence and offering a roadmap for regulation and the worldwide innovation in the field of artificial intelligence. This focus on the human dimension marks a sensitive difference in the global landscape especially if compared with the Russian and Chinese case, in particular. In fact, the EU's "human-centric" approach contrasts sharply with China's focus on economic growth and international competitiveness, as exemplified by China's Social Credit System (Benedetto A., 2023, p.10). This contrast also emerges if it is compared to Russia where AI is strongly used for strengthening the State's control over citizens and political dissent. However, vague

definitions, and a hard coexistence between too different goals risk weakening the efficacy of the AI Act.

These are the main steps of the thesis. In the first chapter I will analyze the origin, structure and architecture of the incoming AI Act in order to understand its significance in the framework of the EU policies and in the light of its impact on the worldwide market. In this section the AI Act emerges as a hybrid document that mixes three different approaches: a values-based approach derived from the AI ethics, a rights-based approach (as a continuation of the GDPR), and a traditional risk-based approach. This chapter analyzes how the AI Act positions the EU as a global leader in AI regulation, leveraging its significant market influence to set ethical standards while balancing innovation and fundamental rights. In the Second chapter I will analyze the comparable regulation of artificial intelligence in Russia. In this framework emerges that the exploitation of legal loopholes under the guise of national security, the need of repressing the political dissent, and the goal of building an effective system of general control largely influence the use of artificial intelligence systems in the country. In the third chapter the case of China is considered. The analysis covers key initiatives such as "Made in China 2025" and the Data Security Law, examining how China navigates ethical considerations amidst critiques of extensive surveillance and control mechanisms. This chapter also explores China's international AI strategies, including cyber-espionage and influence operations, and their implications for global security and democratic integrity.

The thesis concludes by synthesizing findings across chapters, offering comparative insights into how democratic and authoritarian regimes regulate AI and their broader societal impacts. It discusses the implications for global AI governance, highlighting potential threats to individual liberties and opportunities for ethical AI development.

Chapter I:

The Rising Regulation of the Artificial Intelligence in the Global Landscape

1.1 Introduction

The European Union's adoption of the AI Act marks a significant advancement in the regulation of artificial intelligence (AI) technologies across its Member States. Rooted in a complex legal framework encompassing primary legislation, secondary legislation made up of sector-specific regulations, soft law instruments and ethical guidelines, the Act aims to cope with the multifaceted challenges posed by AI's rapid evolution. In order to understand the importance of the European regulation of artificial intelligence in the global landscape, it is thus crucial to analyze its legal underpinnings, regulatory scope, enforcement mechanisms, and overarching objectives in the light of its global scope, namely to analyze its inner architecture considering its interrelation and its effects at the global level.

The EU AI Act emerges as a pivotal response to the expansive impact of AI on diverse domains such as health, safety, digital markets, and other norms anticipating any regulatory initiative of any other country in the world. This fact has given the EU a leading role in the law making in artificial intelligence transforming the EU into the “king maker” of AI regulation of the worldwide landscape. The AI Act consolidates various legislative efforts under a unified strategy aimed at promoting EU values and, in particular, at safeguarding the internal market. While historically, EU regulations focused on specific sectors, the AI Act represents a comprehensive approach to regulate the broad spectrum of AI applications that try to combine trustworthiness of artificial intelligence, protection of fundamental rights and the development of the AI sector in Europe. It mainly mixes three different approaches: a values-based approach derived from the AI ethics (AI High-Level Expert Group 2018), a rights-based approach derived, *inter alia*, by the GDPR (Ruggiu, 2018) and a traditional risk-based approach, derived from the techno-scientific domain. Paradoxically the main novelty has come from this latter domain that has given the AI Act a great visibility at the worldwide level. By categorizing AI systems based on risk levels—from prohibited AI systems posing unacceptable risks to minimal-risk models—the Act tailors regulatory measures to mitigate potential harms while fostering innovation.

Central to the Act's architecture is its commitment to upholding the AI ethical principles and fundamental rights. Drawing from the “Ethics Guidelines for Trustworthy AI” [1] the AI Act implements the ethical principles aimed at achieving the trustworthiness that are strategic in the development of the artificial intelligence at both industrial and societal level (such as respect for human autonomy, prevention of harm, fairness and explicability). Drawing from the EU Charter of Fundamental Rights, the Act prioritizes human dignity, privacy, non-discrimination, and autonomy in AI deployment protecting democracy, the Rule of law and the environment). In this last regard, it aims at establishing a robust enforcement framework involving EU bodies like the AI Office and national competent regulators to ensure consistent compliance and oversight. This regulatory structure not only aims to enhance trust in AI technologies but also positions the EU as a global leader in setting ethical standards for AI governance. However, this acquisition of a global leadership owing to its ability of using the relevance of its internal market and its ability of anticipating the other players in the regulatory field is not without any shortcoming that can affect the effectiveness of the AI Act in the future. In particular, the difficulty of conjugating such different objectives and values has led to a given normative vagueness that cannot but affect the applicability of its provisions. In this context, notwithstanding its limits, the risk-based approach is the main result of this valuable effort of leading the market in the field of artificial intelligence in Europe and abroad.

In the subsequent sections, this chapter delves deeper into the classification system outlined in the AI Act, examines its operational implications for stakeholders, and critically evaluates its strengths and weaknesses in addressing the complex landscape of AI governance. By contextualizing the EU AI Act within its historical narrative and examining its regulatory mechanisms, this study aims to provide a comprehensive understanding of its significance in shaping the future of AI regulation and governance both regionally and globally.

In 2018, the European Commission emphasized the critical importance of regulating how AI technologies are deployed, by whom, and under what circumstances, recognizing that "the way we approach AI will define the world we live in[2] ". Despite the EU's fragmented political landscape and recent challenges like Brexit, the successes of regulations like the GDPR and

[1] AI High-Level Expert Group set up by the European Commission, "Ethics Guidelines for Trustworthy AI" High-Level Expert Group on Artificial Intelligence, European Commission, Brussels, June 2018.

[2] Communication from the Commission on Artificial Intelligence for Europe, 2018, p.1.

the AI Act illustrate the EU's capacity to wield substantial influence as a legislator in digital realms. These regulations are pivotal in shaping global standards for digital technologies and AI development, marking strides towards achieving digital sovereignty. However, it is essential to acknowledge that major players in AI innovation, such as the United States of America, South Korea, and China, wield significant influence globally. This means that the real leaders in this field are not European. While the EU can foster its digital market with several wide ranging policies, its impact on the broader international landscape remains circumscribed by the dominance of these technological powerhouses. Rules, instead, seem to be the real asset of the EU. In this sense, the Ai Act just followed the lesson of the GDPR by using the critical mass of the European market for establishing standards that are able to cross the main limited borders of the European continent. Thus, whilst the innovation in the sector of artificial intelligence is made elsewhere, Europe fixes the rules that are able to impress the direction of the worldwide market.

Initially, this study will explore the origins of the EU AI Act, covering its historical development and legal foundations. Then, a thorough analysis will be conducted to clarify the structural framework and goals outlined within the legislative framework. Additionally, the examination will extend to identifying perceived threats and the corresponding defensive measures devised to uphold a human-centric approach. Concluding this chapter, focus will be on the ongoing discussion regarding the identification of gaps within the regulatory framework in terms of violation or underdevelopment in the protection of the rights and liberties of the European citizens raise by both scholars and human rights and civil liberties watchdogs.

1.2 *Basis for the development of EU AI Act*

The EU AI Act has a wide range of legal acts, treaties, regulations, recommendations and guidelines etc. under its legal basis that highlight its significance within the EU’s framework. For reconstructing this significance it is thus necessary to consider a wide legislative spectrum. Navigating the complex landscape of regulations, particularly when they intersect various domains like health and safety, data protection, and digital markets, poses significant challenges due to the different nature and aim of the concerned acts. This section seeks to address this complexity by categorizing regulations into two distinct groups: primary legislation, secondary legislation that includes sector-specific legislation, ethical guidelines and any other soft law tool. The secondary legislation forms the framework within the action

of the AI Act must be seen in conjunction with any other strategic policy in the field of innovation (like digital sovereignty), whilst the primary legislation establishes the objectives that must be put at the center of the EU's policies in any field (such as artificial intelligence). In this regard, the AI Act appears as a piece of the broader puzzle aimed at building the Eu's digital sovereignty (European data protection, digital market, digital platforms, digital services and artificial intelligence) within the framework of the objectives that are established in its treaties. (the Treaty of the European Union, the Treaty of Functioning of the European Union, the Charter of Fundamental Rights). Each piece of legislation, such as the AI Act, should be thus considered as a part of one strategy, which allows the European Union to achieve its goal of building the bases of its sovereignty in the digital fields. In other words, like GDPR, the Digital Single Market Directive, and the other incoming legislations (the Draft Digital Services Act, the Draft Digital Markets Act, the Draft Data Governance Act) aimed at imposing a single regulation on the European market in the digital field, the EU has regulated its own market with regard to artificial intelligence in order to overcome regulatory barriers within the countries and create an unique homogeneous environment of rules able to foster the innovation in this sector. In this sense, in its strategy on digital sovereignty the EU is using the power that it has from the extent and the richness of its market (comparable to that of the USA or China) for strengthening its rules and in this way its position at the global level. This is in line with its own history. In fact the European Union arose mainly merely with an economic mission, the protection of its internal market, since after the end of the Second World War, European countries believed that only ensuring the safety of their economic and financial exchanges was possible to grant peace in Europe. The protection of the internal market is thus for historical and legal reasons an objective with a fundamental importance within the European Union and this can be seen in its primary legislation, its founding treaties. Even its political dimension, the protection of the fundamental rights of the European citizens (that arose lately owing to the jurisprudence of the Court of Justice), is normally in competition with, and often subordinated to, the goal of the internal market. In this regard, also the EU AI Act has pursued the main purpose of protecting the EU market by laying down uniform provisions for all EU countries and avoiding that market actors could take advantage from differences in legislation of artificial intelligence. While previous legislation has been straightforward and focused on specific fields, the EU AI Act aims to regulate the wide range of areas where rapidly emerging artificial intelligence impacts various spheres, such as personal rights, market regulation, and more.

AI Act thus takes influence from the Primary Legislation first, with regard to its own objectives (the internal market and the ethical principles and fundamental rights that are at the basis of the European Union), and from the Secondary Legislation with regard to the converging policies aimed at building the digital sovereignty of the European Union). Within the framework of the Secondary Legislation we have also to consider Specific Sectorial Legislations (for example in the field of health or machinery) and Guidelines laid down for helping the construction of a governance framework in the field of artificial intelligence that aligns with the founding ethical principles of the European Union and the main ethics debate in the field of artificial intelligence.
The scheme of AI Act is thus the following:

Primary Legislation: This category includes foundational legal documents that directly influence the EU AI Act. It encompasses the treaties that establish the objectives for EU policies (such as the AI governance). These include:
- the Charter of Fundamental Rights of the European Union[3]
- the Treaty on the Functioning of the European Union (TFEU)[4]
- the Treaty on European Union (TEU)[5]

Secondary Legislation: These are regulations, directives, and acts that provide specific guidelines, requirements, and procedures that enact the policies of the European Union, also in the field of AI governance. They may focus on data protection, consumer rights, digital services, etc. These include:
- the General Data Protection Regulation (GDPR)[6]
- the Digital Single Market Directive (DSM Directive)[7]

[3] European Union, Charter of Fundamental Rights of the European Union, adopted on October 2, 2000 and entered into force, after the Lisbon Treaty (2007), on December 1, 2009.

[4] European Union, Treaty on the Functioning of the European Union, originally signed on March 25, 1957, as the Treaty of Rome, and subsequently amended by the Treaty of Lisbon, which entered into force on December 1, 2009.

[5] European Union, Treaty on European Union, originally signed on February 7, 1992, in Maastricht, and subsequently amended by the Treaty of Lisbon, which entered into force on December 1, 2009.

[6] European Union, General Data Protection Regulation (GDPR), Regulation (EU) 2016/679 of the European Parliament and of the Council, adopted on April 27, 2016, and applicable from May 25, 2018.

[7] European Union, Digital Single Market Directive (DSM Directive), Directive (EU) 2019/790 of the European Parliament and of the Council, adopted on April 17, 2019, and applicable from June 7, 2021.

- the Draft Digital Services Act (DSA)[8]
- the Draft Digital Markets Act (DMA)[9]
- the Draft Data Governance Act[10]

Sector-Specific Legislation: This category encompasses regulations tailored to specific sectors or industries impacted by AI, such as healthcare, transportation, finance, etc. These regulations often address sector-specific concerns related to safety, privacy, and ethical use of AI. These include:
- the Regulation of Union health and safety harmonization legislation[11]
- the Machinery Regulation (for AI-integrated machinery)[12]
- the Consumer protection laws[13]

Ethical Guidelines and Frameworks: This category includes non-binding ethical guidelines and frameworks developed by expert groups, commissions, or international organizations. These guidelines provide ethical principles and standards for the development and deployment of AI technologies. These include:
- the Ethical guidelines for trustworthy AI by the AI High-Level Expert Group[14]

---

[8] European Union, Draft Digital Services Act (DSA), Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (COM/2020/825 final), introduced on December 15, 2020.

[9] European Union, Draft Digital Markets Act (DMA), Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (COM/2020/842 final), introduced on December 15, 2020.

[10] European Union, Draft Data Governance Act, Proposal for a Regulation of the European Parliament and of the Council on European data governance (COM/2020/767 final), introduced on November 25, 2020.

[11] European Union, Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, Official Journal of the European Union, L 169/1, 25 June 2019.

[12] European Union, Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, Official Journal of the European Union, L 169/1, 25 June 2019.

[13] European Union, Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council, Official Journal of the European Union, L 304/64, 22 November 2011.

[14] AI High-Level Expert Group set up by the European Commission, "Ethics Guidelines for Trustworthy AI," High-Level Expert Group on Artificial Intelligence, European Commission, Brussels, June 2018.

- the European Commission's White Paper on Artificial Intelligence[15]

In the following sections, secondary legislation, sector-specific legislation and ethical guidelines and framework will be reviewed with estimation of the effect on the EU AI Act.

The Charter of Fundamental Rights of the European Union is applicable to the EU AI Act and also to almost all legislative initiatives and common projects. Principles and rules established in the EU Charter provide the core values to ensure that legislative decisions in all spheres align with the fundamental rights and values described therein. Its aim is to ensure the protection of fundamental rights and freedoms across the European Union, particularly the personal freedom, freedom of thought, non-discrimination, the rights to privacy and data protection, fair and transparent decision-making, which may be at stake when artificial intelligence systems are used.

For example, the Charter of Fundamental Rights ensures that AI systems do not violate the right to non-discrimination through biases or other automated decisions. The rights to privacy and data protection ensure users do not have to worry about misuse of personal data or unauthorized access by AI systems. Ultimately, the principles outlined in the Charter provide a guiding framework for the development and regulation of new technologies, promoting responsible and ethical development that respects human dignity, autonomy, and privacy.

Additionally, the European Convention on Human Rights (ECHR)[16] and the jurisprudence of the Strasbourg Court play a significant role in influencing the application of certain rights, particularly personal freedom, civil and political freedoms, the respect of democracy and the basic principles of the Rule of law, privacy and data protection, thereby strengthening their protection in Europe. Although the ECHR rights are encompassed within the Charter and the Charter addresses contemporary issues not covered by the Convention (such as human cloning and data protection), the influence of the ECHR and the Strasbourg Court's interpretations still enhance and complement the protection offered by the Charter. Therefore, both the Charter and the ECHR collectively provide a robust framework for ensuring that AI technologies respect and uphold fundamental human rights.

[15] European Commission, "White Paper on Artificial Intelligence - A European approach to excellence and trust," COM(2020) 65 final, Brussels, 19 February 2020.
[16] Council of Europe, European Convention on Human Rights and Fundamental Freedoms (ECHR), signed in 1050 and entered into force in 1953.

Treaty on the Functioning of the European Union and Treaty on European Union

Some articles from the Treaty on the Functioning of the European Union and Treaty on European Union also serve as a legal foundation of EU AI Act.
TFEU Article 16 enshrines the right to the protection of personal data for all individuals within the EU. It provides a fundamental basis for ensuring that AI systems respect individuals' privacy rights and data protection principles, which are crucial components of the EU AI Act.

TFEU Article 114 outlines the procedures for adopting measures to achieve the objectives of the internal market. It enables the harmonization of laws among Member States to ensure the proper functioning of the internal market. In the context of the EU AI Act, Article 114 provides the legal basis for harmonizing AI regulations to promote consistency and coherence in the EU's approach to AI governance. This focus on the internal market is crucial. The EU arose as an economic organization, and its main mission has always been the protection of the internal market. This objective led to the adoption of the GDPR first and now the AI Act. Pursuing this objective of uniform regulation of the internal market, whether for data through the GDPR or for artificial intelligence through the AI Act, the EU has also protected the fundamental rights of European citizens.

TFEU Article 294 outlines the legislative procedures for adopting Union acts and other legal provisions. It ensures transparency, accountability, and democratic legitimacy in the decision-making process.

TEU Article 2 sets forth the core values of the EU, which encompass respect for human dignity, freedom, democracy, equality, and the Rule of law. These values underpin the development and implementation of the EU AI Act, ensuring that AI technologies are developed and used in a manner that upholds fundamental rights and democratic principles. Article 3 establishes what can be considered the main objective of the European Union that is the internal market to which the freedoms of circulation of persons and goods (including AI systems and services) are strictly linked. It is for ensuring a uniform regulation of the digital technologies that both GDPR and now the Ai Act have been adopted. All the EU values are balanced with this objective.

With regard to the Secondary Legislation, in order to summarize information we can categorize the legal act in terms of impact on the EU AI Act.

Particularly important are those acts that enforce the EU strategy aimed at construing an EU digital sovereignty starting from the GDPR.
Digital Single Market Directive (DSM Directive) was published in 2017 and edited in May 2019, with the aim of harmonizing EU copyright law and adjusting established rules to adapt to the changing digital landscape. This Directive establishes a united digital market in the European Union without trade barriers and ensuring free competition in cyberspace. The DSM directive is crucial for EU AI Act since it is one of the legislative pieces which lies at the core of the new AI regulations. DSM addresses free flow of non-personal data, cybersecurity and protection aspects. Moreover, rules which are relevant to digital services, data protection and online commerce create a specific framework within businesses in the digital sphere and they have to consider those rules during the development of the new technologies, including AI.

The General Data Protection Regulation (GDPR) adopted in 2016 and entered force on 25th May 2018 with the specific purpose of ensuring an uniform regulation for the circulation of data in Europe and protecting the personal data of the European citizens. If the DSM Directive is focused mostly on the regulation of non-personal data, GDPR is directly engaged in this field by ensuring privacy and protection of personal data. GDPR was created to replace the previous version of the Data Protection Directive (1995) which was outdated, allowed different legislations on privacy among Member States, and could not cover new situations that emerged with technological progress. Beyond a set of rights strengthening the privacy of the European citizens, one of the most important innovations emerged with GDPR is safe design by default (the principle of privacy by design), which means that security measures have to be established and ensured to comply with existing regulations on the development stage of a given technology. Another important feature of this law is that decision-making is based on the automized proceeding, including profiling. It is done for people to understand the mechanical processing of the technologies which make them more trustworthy. Individuals have a right to receive information about the logic behind automated decisions, to dispute decisions that significantly impact them and not to be subjects solely to automated decision-making, except certain situations. It is essential because during profiling automated decision-making can perform biases and to avoid it we should ensure the

possibility of human decision making. By these two important implications GDPR impacts EU AI Act significantly. The main problem which is tightly interconnected with AI is the black box problem, which basically means that even the developer of these technologies sometimes cannot tell how the decision was made, because the system is self trained and cannot fully explain why the decision is made like that. By closely monitoring GDPR's application in the context of AI, GDPR principles help to ensure that AI systems do not infringe individuals' rights and that it is possible to detect that decision was made by machine and how it was made.

To illustrate how the DSM Directive and GDPR influenced EU AI Act: by categorizing data on personal and non personal both legislative acts ensure both data protection rights and cybersecurity but at the same time it provides AI with non personal data so that systems can be trained and developed without harm to others. This example shows that the European Union emphasizes both the development and innovation and personal freedoms and human rights developing a framework which benefits both the public and private sector.

Draft Digital Services Act (DSA):
The DSA aims to regulate digital services, including online platforms, to ensure accountability, transparency, and fairness. It addresses issues such as content moderation, algorithmic transparency, and access to data. The provisions on recommender systems and research data access in the DSA influence how AI systems are deployed and governed within digital services.

Draft Digital Markets Act (DMA):
The DMA focuses on regulating large online platforms to prevent monopolistic practices and ensure fair competition in digital markets. It addresses issues such as platform dominance, interoperability, and self-preferencing. The DMA's regulations concerning AI-relevant hardware, operating systems, and software distribution impact the development and implementation of AI technologies within digital markets. They contribute to creating a level playing field for AI innovation and preventing anti-competitive behavior.

Draft Data Governance Act:
The Draft Data Governance Act facilitates data sharing frameworks while ensuring privacy, security, and fair competition. It aims to promote data sharing for AI development and

innovation. The Data Governance Act's provisions on data sharing frameworks contribute to facilitating access to data for AI development and training purposes. They complement the AI Act by promoting responsible data sharing practices that align with ethical and legal standards.

European Commission’s White Paper on Artificial Intelligence

The White Paper emerged with current technological development to build an established coordinated approach to AI policy within the EU and can be seen as a part of the strategy of protection of the European market with regard to AI. The aim of the paper is to address emerging challenges and opportunities presented by AI, considering also ethical issues, potential risks and competitiveness of the European Union in the global AI landscape. The White Paper proposes regulatory frameworks, investment strategies and ethical guidelines to steer the development and deployment of AI technologies across multiple sectors. Thus, The White Paper addresses specific challenges posed by AI such as transparency, accountability, fairness and safety.

Among all the documents listed above it seems that White Paper influences the EU AI Act the most as it provides the strategy of the future development of AI building framework within which it can be developed without harm to society. By outlining key policy objectives and principles, the White Paper helps shape the content and scope of the EU AI Act, ensuring that it reflects Europe's vision for responsible and human-centric AI.

Ethical guidelines for trustworth AI

The AI High-Level Expert Group (AI HLEG) released ethical guidelines for trustworthy AI in 2019. These guidelines aim to ensure that AI systems are developed and used in a manner that is ethical, transparent, and respects fundamental rights and values. The guidelines outline seven key requirements for trustworthy AI, which include ethical principles such as beneficences (do the well) and non-maleficence (do not harm), fairness, transparency, accountability, and respect for human autonomy. They also emphasize the importance of human oversight, robustness and safety, privacy and data governance, as well as non-discrimination and societal and environmental well-being.

This document has great importance for the AI Act since, like the GDPR, it has significantly influenced the current shape of the AI Act leading to follow a more traditional risk-based

approach of techno-scientific origin (that can be reflected in the non-maleficence principle). While the GDPR is mainly aimed at implementing a right (the right to data protection), the AI Act adds to several fundamental rights (derived from the EU Charter) a set of ethical principles that stem from the ethical debate on artificial intelligence, whose the risk-based approach can be seen as a reflection. Thus, the AI Act is a hybrid regulatory framework that mixes a rights-based approach and a risk-based approach that lead to establish several levels of risks for artificial intelligence.

The EU AI Act aims to translate these ethical principles into legally binding regulations since their adoption is seen as strategic for building the trustworthiness necessary for the diffusion of AI systems. By aligning the provisions of the AI Act with the ethical guidelines, the EU seeks to establish a regulatory framework that promotes the development and use of AI systems in a manner that upholds these ethical principles. This means that the regulations outlined in the AI Act will likely reflect the values and objectives set forth in the ethical guidelines, thereby ensuring that AI technologies deployed within the European Union adhere to principles of human-centric AI, beneficence, non-maleficence (risk-based approach), fairness, transparency, accountability, and societal well-being.

Beside the most important legal acts which can be considered the core foundation of EU AI Act we can consider several other legislations, which influenced the EU AI Act.

Draft Machinery Regulation:
The draft Machinery Regulation revises the Machinery Directive to address AI-related challenges in health and safety standards for machinery. It aims to ensure that AI-integrated machinery meets safety requirements and does not pose risks to workers or consumers.
The Machinery Regulation's provisions on AI-related health and safety standards complement the AI Act by ensuring that AI-integrated machinery complies with safety requirements outlined in the Act. They contribute to mitigating risks associated with AI deployment in industrial settings.

Product Liability Revision:
The revision of the Product Liability Directive aims to establish liability frameworks specifically tailored to AI-related products. It clarifies who is responsible in cases of harm

caused by AI systems and ensures that adequate compensation mechanisms are in place. The revision of the Product Liability Directive enhances the legal framework for addressing liability issues related to AI products. It complements the AI Act by providing clarity on liability standards and ensuring accountability for harm caused by AI systems.

1.2 *The Steps that led to the Current Version*

The EU AI Act can be seen as an effort to establish values and standards for the international community, although it is primarily aimed at protecting the EU market and ensuring the EU plays a significant role on the global stage. While the EU is not a leader in the AI industry and digital technology, it leverages the strength of its vast market and its regulatory power, as exemplified by the GDPR. The historical narrative of the pivotal stages leading to the establishment of the EU AI Act is crucial, as it places the Act within the broader context of AI governance evolution. This contextualization underscores the deliberative process, stakeholder involvement, and policy formulation that shaped the Act. By examining its historical journey, we gain valuable insights into the rationale, objectives, and collaborative efforts that have shaped the Act, providing a deeper understanding of its significance and implications for AI regulation. While protecting the rights of EU citizens is an important outcome, it is often a secondary goal to the primary objectives of market protection and global influence.

In October 2017, the European Council emphasized the need for AI regulation to ensure high levels of data protection, digital rights, and ethical standards. The European Parliament, together with the European Economic and Social Committee, provided extensive recommendations. Subsequently, the European Commission began developing a European approach to AI. The initiative aimed to enhance technological potential and the integration of AI both in the public and private sectors: preparing for socio-economic changes due to modernization, ensuring an ethical and legal framework based on Union values and the corresponding Charter of Fundamental Rights of the EU by compiling guidance on existing product liability rules, conducting a detailed analysis of emerging issues, and collaborating with stakeholders within the European AI Alliance to develop guiding principles for AI ethics (European Commission, 2018).

On 10th April, 2018, 24 member states signed a Declaration of Cooperation, agreeing to work together on a coordinated AI plan. The main goals of this declaration will be to maximize the impact of investments at the EU and national levels, encourage synergy and cooperation within the EU, share best practices, and jointly determine the way forward to ensure that the EU as a whole can compete on a global scale.

To support the declaration, the Commission has augmented its investment in artificial intelligence to €1.5 billion through the Horizon 2020 program and allocated an additional €2.5 billion to foster collaboration between public and private sectors. In the press release from 25th April 2018, commission stated that these investments are intended to ensure the interest and involvement of research centers in testing and innovating AI, as well as to strengthen the dissemination of technologies in various sectors, including small and medium-sized enterprises. Additional measures were also introduced to encourage research and innovation in transitioning AI from laboratories to the market.

Moreover, The European Commission initiated a process of consultation and engagement with stakeholders, including industry representatives, civil society organizations, academia, and citizens. Through public consultations, expert workshops, and stakeholder dialogues, the Commission gathered input and feedback on the opportunities and challenges associated with AI, as well as potential regulatory approaches.

On 25th April 2018, the European Commission issued a communication to the European Parliament proposing the establishment of a coordinated approach to AI to address challenges and seize opportunities arising from new technological advancements. Through regulation, the European Commission aims to ensure that Europe is competitive in the AI landscape, that the unintended consequences of modernization (unemployment and development gaps) will be offset by inclusion policies and providing equal opportunities, and that the approach to dealing with AI is interconnected with the core values of the European Union principles.

In February 2020, the European Commission published the White Paper on Artificial Intelligence - A European approach to excellence and trust. The White Paper outlined key policy areas and proposals for regulatory frameworks, investment strategies, and ethical guidelines to guide the development and deployment of AI technologies in the EU. It served as a roadmap for shaping EU AI policies and initiatives.

Building on the recommendations outlined in the White Paper, the European Commission developed a legislative proposal for the EU AI Act. This proposal underwent rigorous impact assessments to evaluate its potential economic, social, and environmental implications. The Commission also conducted consultations with EU member states, the European Parliament, and other stakeholders to gather feedback on the proposed legislation.

On 9th December 2023 the legislative proposal for the EU AI Act was presented to the European Parliament and the Council of the European Union for negotiation and adoption. Intensive discussions and negotiations took place to reach agreement on the content and scope of the legislation. The legislative process involved multiple rounds of revisions and amendments to address concerns and incorporate feedback from stakeholders.

On 13th March 2024 the EU Parliament approved the AI Act, and then it was formally adopted by the Council on 21 May 2024.
After reaching consensus among EU institutions, the EU AI Act is waiting to be published in the Official Journal of the European Union for being enacted into law. Once entered into force, Member states are responsible for implementing the provisions of the Act within their national legal frameworks, ensuring consistency and harmonization of AI regulations across the EU.

1.3 *Architecture and Objectives of the AI Act*

The EU AI Act is structured around key purposes and principles, focusing on protecting the internal market and safeguarding fundamental rights. This section examines the scope, purpose, activities covered, enforcement mechanisms, prohibitions, and the threats and opportunities addressed by the Act. The scope includes all actors in the AI supply chain, covering AI systems used within the EU and applying extraterritorially. The purpose is to enhance the internal market and encourage AI adoption that prioritizes human well-being and reliability. It ensures protection for health, safety, and fundamental rights, promoting responsible AI development through harmonized rules and support for innovation. Enforcement involves EU bodies like the AI Office, AI Board, Scientific Panel, and Advisory Forum, alongside national competent regulators. Coordination between EU and national regulators ensures consistent enforcement. Key principles include protecting the internal

market with uniform regulations, safeguarding fundamental rights like non-discrimination, privacy, and data protection, and emphasizing trustworthiness through ethical principles such as fairness, transparency, accountability, human oversight, robustness, and safety.

Purpose

Chapter 1, Article 1 of the EU AI Act outlines its fundamental purpose, which is to enhance the functioning of the internal market and encourage the adoption of artificial intelligence (AI) that prioritizes human well-being and reliability. This legislation aims to guarantee an advanced level of protection for health, safety, and fundamental rights, as outlined in the Charter of Fundamental Rights of the European Union. By mitigating the adverse effects of AI systems and promoting innovation, the Act seeks to develop a regulatory framework that promotes the responsible development and deployment of AI technologies across the Union. Additionally, it establishes harmonized rules for the introduction, operation, and utilization of AI systems, including provisions for prohibiting certain AI practices, specifying requirements for high-risk AI systems, and imposing obligations on operators. Furthermore, the Act includes transparent guidelines for certain AI systems, rules for the introduction of general-purpose AI models to the market, and provisions for market monitoring, surveillance governance, and enforcement. It also includes measures to support innovation, with a particular focus on small and medium-sized enterprises (SMEs), including startups.

The EU AI Act implements a risk-based approach to regulate artificial intelligence (AI) systems, aiming to mitigate potential harm while promoting innovation. Chapter 1, Article 5 of the Act delineates specific prohibitions and classifications for different types of AI systems.

Scope of the AI Act

In chapter 1 Article 2 the EU AI Act applies comprehensively across the entire AI supply chain, encompassing various actors involved in the development, marketing, and use of AI systems within the Union. These actors include:

1. Providers: Developers of AI systems or general-purpose AI models who place these systems on the market or put them into service under their own name or brand.

2. Deployers: Entities or individuals who use AI systems, including customers or providers.

3. Importers: Individuals who place AI systems on the EU market.

4. Distributors: Entities in the supply chain, other than providers or importers, who make AI systems available on the EU market.

5. Operators: Designate providers, product manufacturers, deployers, authorized representatives, importers, and distributors.

6. Affected Persons: Those potentially impacted by AI systems' use.

Activities Covered and Exclusions

The AI Act covers various activities related to AI development, placement on the market, and use within the EU. It applies to developers placing AI systems or general-purpose AI models on the EU market, as well as to deployers established in the EU. Additionally, it applies extraterritorially to providers and deployers outside the EU if their AI system's output is used within the EU.

Exclusions from the AI Act include AI used for military, defense, or national security purposes, AI used by public authorities outside the EU (with international agreements in place), and AI systems developed for scientific research and development. Furthermore, personal, non-professional activities involving AI are exempt, along with AI systems released under free and open-source licenses, except under certain conditions concerning prohibited AI, high-risk AI, and GPAI models.

1.4 Analysis of the threats outlined in the EU AI Act and measures of protection

Classification

The architecture of the EU AI Act is organized around protecting the internal market and fundamental rights. The Act ensures that AI systems deployed within the EU adhere to principles of trustworthiness, transparency, accountability, and societal well-being. By implementing a risk-based approach, the Act aligns with the EU's core values of protecting human dignity, autonomy, privacy, and non-discrimination, as outlined in the EU Charter of

Fundamental Rights. This regulatory framework promotes the responsible development and deployment of AI technologies, balancing innovation with the protection of fundamental rights, thereby fostering trust in AI systems across the European Union. This approach acknowledges that not all AI systems pose the same level of risk and thus require tailored regulations. By categorizing AI systems based on their risk levels, the Act aims to implement proportionate measures to mitigate potential harm while fostering innovation. In the following overview, we will delve into the classification system outlined in the Act, highlighting the key requirements and prohibitions for each category.

The Act categorizes AI systems into four distinct groups based on risk levels (introduced in Titles 2, 3, and 5). There are four levels of risk outlined in the Act: unacceptable risk, high-risk, limited risk and minimal risk.

1. Prohibited AI Systems (Unacceptable risk): These include AI systems that utilize subliminal techniques to manipulate behavior, exploit vulnerabilities based on age or disability, or engage in social scoring and biometric categorization without targeted consent. Such systems are banned in the EU as they pose unacceptable risks to individuals and violate fundamental rights.

2. High-risk AI: Examples of high-risk AI systems include those used as safety components in critical fields such as medical devices, autonomous vehicles, or infrastructure monitoring. These systems have the potential to cause significant harm if they malfunction or are used improperly, so most of the text address regulations for this type of system.

3. Limited risk AI Models: Limited AI models encompass systems with specific transparency requirements or those designed for narrow applications. Examples include chatbots programmed to disclose their non-human nature to users or emotional recognition systems used for targeted interventions in mental health care.

4. Minimal risk AI Models: Minimal risk AI models are systems that pose the least threat to health, safety, and fundamental rights, and thus, they are subject to minimal regulatory oversight. These models are typically unregulated under the EU AI Act due to their limited impact on individuals and society. Among minimal risk AI Models the EU AI Act mentions

AI-enabled video games and AI systems are designed to identify and filter out unwanted emails.

Activities Falling Under the Category of "Prohibited AI Systems"

Under Article 5, certain AI practices are prohibited due to their potential harm or infringement on fundamental rights. These include:

1. Subliminal Techniques: Banning AI systems from using deceptive techniques to manipulate behavior or cause harm.

2. Exploitation of Vulnerabilities: Prohibiting the exploitation of vulnerabilities based on age, disability, or social/economic status.

3. Biometric Categorization: Preventing AI systems from inferring sensitive personal attributes like race, religion, or sexual orientation through biometric categorization.

4. Unfair Classification: Prohibiting the evaluation or classification of individuals based on social behavior or personal characteristics in ways that lead to unfair treatment. In this category we can add such activities as social scoring and assessing the risk of individuals committing criminal offenses.

5. Real-time Remote Biometric Identification: Regulating the use of real-time biometric identification systems for law enforcement purposes, allowing strict usage guidelines only for specific purposes such as finding missing persons or preventing serious threats. As an example, compiling facial recognition databases and emotional inferring are acceptable only if it is used for safety and medical reasons. For example it can be used to search for the missing person, prevent life threat and identify suspicious crimes, but requires several compliance procedures. Among them, prior to deployment, law enforcement agencies are required to conduct a fundamental rights impact assessment and register the system in the EU database. However, in urgent situations, deployment may proceed without prior registration, as long as it is promptly registered afterward. Additionally, authorization from a judicial or independent administrative authority is necessary before deployment, although in cases of urgency, deployment can occur initially without authorization, provided that authorization is

sought within 24 hours. If authorization is denied, deployment must cease immediately, with all associated data, results, and outputs deleted.

Systems Falling Under the Category of the "High-risk"

The EU AI Act introduces stringent regulations for high-risk AI systems under the chapter 3, categorizing them into two main categories:

Category 1: This includes AI systems that serve as safety components in products subject to EU sectoral legislation on product regulation. These systems are required to undergo third-party conformity assessments to ensure compliance with safety standards. For example, self-driving cars. The AI system controlling the car's safety features, like collision detection and emergency braking, falls into this category. It needs to undergo assessments to ensure it meets safety standards set by the EU for automotive products.

Category 2: Includes AI systems that present notable dangers to the well-being, security, or basic rights of individuals. However, certain AI systems are excluded from high-risk classification if they do not pose a significant risk of harm or materially influence decision-making outcomes, as determined by a Fundamental Rights Impact Assessment. Here, the example can be a medical diagnostic AI system used in hospitals. This system analyzes patient data to help doctors make diagnoses. Since it directly affects people's health and safety, it falls into this category. However, a medical appointment scheduling AI system used in hospitals might not be classified as high-risk because it doesn't directly impact patient health or safety.

For high-risk AI systems, the EU AI Act mandates strict compliance obligations in various areas.

- Assessment and Registration: High-risk AI systems must undergo comprehensive risk assessments and be registered with regulatory authorities.

- Risk Management and Human Oversight: Providers of high-risk AI systems are required to implement robust risk management processes and ensure human oversight to mitigate potential harms.

- Data Governance and Security: Stringent data governance and security measures must be implemented to protect the privacy and security of individuals' data.

- Technical Documentation, Record Keeping, and Transparency: Providers must maintain detailed technical documentation, records, and transparency measures to ensure accountability and facilitate regulatory oversight.

The Act also implements a two-tiered approach for general-purpose AI models, with fewer obligations for models deemed to have no systemic risk and higher obligations for models with systemic risk. Classification of systems of risk is based on the evaluation of high-impact capabilities using appropriate technical tools and methodologies. AI systems are always considered high risk if they perform profiling and providers which consider their systems high risk must document it before placing it on the market.

General purpose AI Models

AI systems can be designed not only with specific purposes, but also with a general-purpose that can have several different applications in daily life (e.g. chatbots, ad generation, decision assistants, spambots, translation, etc.). General purpose AI systems are increasingly used for powerful applications in medicine and healthcare, finance, life sciences and chemistry, and lately, even in programming and machine learning. General Purpose AI (GPAI) refers to AI models that are designed to be adaptable and versatile, capable of performing a wide range of tasks across different applications. These models are not specifically tailored to a single use case but rather have the flexibility to be deployed in various scenarios. They are often trained on large datasets and are intended to generate outputs that can be applied across different domains. These AI models are designed to perform a wide range of tasks and can generate content autonomously. Examples include natural language processing models used for text generation, image generation, or creative applications like art and music generation. The most famous examples of General Purpose AI systems areAlphaStar, Chinchilla, Codex, DALL•E 2, Gopher, GPT-4, MuZero, PaLM and Wu Dao 2.0. Some of these systems have already caused alarm by propagating extremist content, exhibiting anti-Muslim bias or inadvertently revealing personal data (Future of Life institute, 2022, p. 4).

The European Union's AI Act outlines GPAI separately in Chapter 5 because these models have unique characteristics and potential risks that warrant specific attention. While GPAI

models may not pose the same level of risk as high-risk AI systems, they still have the potential to impact individuals' rights, safety, and well-being. Therefore, it's essential to establish regulations and guidelines tailored to the characteristics and potential risks associated with GPAI.

General Purpose AI (GPAI) models are not inherently categorized as limited or minimal risk; rather, their risk level depends on factors such as their application, impact, and potential systemic risks they may pose. Limited Risk GPAI models typically have minimal potential harm and are used in applications that do not significantly impact health, safety, or fundamental rights, such as AI-enabled video games or spam filters. On the other hand, High-Risk GPAI models present significant potential risks due to their application in critical sectors or their potential to affect fundamental rights and safety, such as AI systems used in medical diagnostics or biometric identification. The EU AI Act outlines regulations for both limited and high-risk GPAI models, ensuring compliance with transparency requirements and imposing stringent measures for high-risk models to mitigate potential risks and ensure safety and reliability.

The obligations for General Purpose AI (GPAI) model providers involve several key requirements under the EU AI Act. Firstly, all providers of GPAI models are mandated to furnish comprehensive technical documentation, user instructions, and ensure compliance with copyright laws outlined in the Copyright Directive. Additionally, they are required to disclose a summary detailing the content utilized for training the AI models. For providers offering free and open license GPAI models, adherence to copyright regulations and publication of a training data summary suffice, unless these models pose systemic risks. In cases where GPAI models present systemic risks, regardless of their licensing status, providers must undertake rigorous model evaluations, engage in adversarial testing, monitor and report serious incidents, and implement robust cybersecurity measures to safeguard against potential threats.

In addition to the general obligations outlined earlier, providers of General Purpose AI (GPAI) models deemed to pose systemic risks must adhere to additional requirements. These include conducting thorough model evaluations, which involve comprehensive adversarial testing to identify and mitigate systemic risks effectively. Furthermore, they are obligated to assess and address potential sources of systemic risks associated with their AI models.

Additionally, they must maintain detailed records of any significant incidents and the corresponding corrective actions taken, promptly reporting them to both the AI Office and the relevant national regulatory authorities. Lastly, these providers are tasked with implementing robust cybersecurity measures to safeguard their AI systems against potential cyber threats.

Enforcement

The enforcement of the EU AI Act involves a two-tier mechanism that encompasses both EU bodies and national competent regulators. At the EU level, several entities are tasked with supervising and enforcing different aspects of the regulation:

1. AI Office: Situated within the EU Commission, the AI Office oversees and enforces the provisions outlined in the chapter concerning general-purpose AI. It plays a crucial role in ensuring compliance and addressing any violations that may arise.

2. AI Board: This body serves to provide guidance and advice on the implementation of the AI Act. It coordinates efforts across EU member states, draws recommendations and opinions, and contributes to the effective enforcement of the regulation.

3. Scientific Panel: Comprising experts in the field of AI, the Scientific Panel is responsible for assessing systemic risks associated with AI technologies. Their evaluations help inform decision-making processes and enhance the regulation's effectiveness in mitigating potential harms.

4. Advisory Forum: This forum brings together technical experts to provide specialized knowledge and insights to the EU Commission. Their input aids in the development of policies, guidelines, and best practices related to AI regulation.

In addition to EU bodies, national competent regulators also play a significant role in enforcing the AI Act within their respective jurisdictions. While the Act provides a framework for regulation at the EU level, member states are responsible for establishing their own regulatory bodies to oversee AI-related matters. For example, Spain has already established its AI regulatory body.

However, the delegation of enforcement authority to member states raises concerns about the possibility of multiple and competing regulators operating at the national level. To mitigate

this risk, coordination and harmonization efforts between EU and national regulators are essential to ensure consistent enforcement and compliance across the Union.

Furthermore, some Data Protection Authorities (DPAs) have positioned themselves as competent regulators in the AI domain, further highlighting the need for clarity and coherence in enforcement mechanisms to avoid conflicts with the EU AI Act.

1.5 *The Debate on the Regulation of the AI*

Following the proposal of a comprehensive regulation of artificial intelligence, several criticisms have been put forward by scholars and civil rights organizations that mainly focused on the loopholes and threats posed to fundamental rights and civil liberties. Drawing from scholars and civil liberties union's critiques, this chapter aims to provide a comprehensive understanding of the challenges and limitations of the EU AI Act in safeguarding fundamental rights and civil liberties in the context of rapidly evolving AI technologies and their societal impacts.

Veale and Borgesius (2021) scrutinize the EU AI Act's scope, particularly concerning restrictions on AI systems and their intentions. They highlight a critical limitation in the Act's provision, which restricts AI systems only to subliminal techniques and intentions to cause harm. This restriction significantly constrains the Act's scope, as it excludes AI systems not explicitly designed to harm individuals. As a result, various issues and gaps in harm regulations arise, potentially leaving out AI applications that could still pose risks to society (Veale & Borgesius, 2021, p. 99). Smuha et al. (2021) supported the criticism stating that it doesn't require any obligations for using subliminal techniques, which might be acceptable if done with the full and informed consent of the affected individuals, and under careful monitoring and strict supervision (p.21).
Wörsdörfer (2024) identifies flaws in the EU's risk-based approach to regulating AI systems, particularly regarding scope and effectiveness. Critics highlight several issues, including insufficient testing and public debate, unclear risk thresholds, and the amendability of high-risk systems lists. One major criticism is that the EU AI Act under-regulates non-high-risk AI systems, despite their potential severe impacts on individuals and society. This limited regulation leaves a wide range of AI systems unregulated, raising concerns about their societal impact and the Act's effectiveness in safeguarding fundamental rights and civil liberties (Wörsdörfer, 2024, p. 106-126; AlgorithmWatch, 2021).

Furthermore, critics point out the reliance on self-assessment and standardization processes in the EU AI Act, which may render the rules on prohibited and high-risk AI practices ineffective. Delegating substantial regulatory power to standardization bodies and notified bodies raises ethical concerns, including a lack of democratic oversight, inadequate stakeholder participation, power asymmetries, lack of transparency, and consensus-finding problems. These criticisms highlight significant shortcomings in the Act's regulatory framework, particularly regarding scope and effectiveness in addressing the complex challenges posed by AI technologies (Wörsdörfer, 2024). In his paper Ho-Dac, M. (2024) argues that the process of creating standards for AI regulation lacks transparency and democratic oversight. Standards are developed by private organizations known as European Standardization Organizations (ESOs), and are not subject to parliamentary debate or public scrutiny like laws in a democracy (Ho-Dac, M., 2024). Instead, they are created through consensus among stakeholders, often dominated by major international companies with vested interests (Ho-Dac, M., 2024). Moreover, these standards are not readily accessible to the public and may require payment to access due to intellectual property rights. This lack of transparency poses several problems. Firstly, it limits critical analysis by academics and makes it challenging for legal practitioners to use them as a source of information in legal cases (Ho-Dac, M., 2024, p.9). Secondly, integrating fundamental rights considerations into standards raises questions about delegating the implementation of fundamental rights to private bodies (Ho-Dac, M., 2024, p.15). Critics argue that private organizations lack the expertise and legal competence to regulate fundamental rights effectively.

Almada and Petit (2023) critique the EU AI Act, arguing that it inadequately addresses human rights issues, primarily due to its focus on product safety rather than ethical considerations. Product safety approach may lead to suboptimal AI regulation in three key ways. First, they inaccurately suggest that AI systems have specific purposes, ignoring the complexity of general-purpose AI (Almada & Petit, 2023, p.7). Second, they focus on safety risks, which doesn't address fundamental rights issues (Almada & Petit, 2023, p.7). Third, they rely too much on existing product safety institutions, overlooking the need for new capabilities to protect rights in AI contexts (Almada & Petit, 2023, p.7). They contend that this approach overlooks critical concerns related to the potential erosion of social trust and public legitimacy by technologies like chatbots and deepfakes (Almada & Petit, 2023, p.20).

By viewing human rights primarily through a product safety lens, the Act fails to fully address the broader ethical implications of AI technologies and their impact on society. Moreover, Almada and Petit (2023) highlight the challenge of predicting the functionalities and potential use cases that might raise safety and ethical concerns, particularly in the context of generative AI such as large language models and chatbots. This uncertainty underscores the need for a more comprehensive and nuanced approach to regulating AI systems, one that considers their potential societal impacts beyond mere product safety (Almada & Petit, 2023).

Castets-Renard & Besse (2022) criticize the Act for prioritizing economic interests over moral concerns and human rights. The EU AI Act primarily emphasizes economic and business priorities over moral concerns and human rights. This focus is evident because the Act aims to boost innovation, competitiveness, and market growth within the AI sector, potentially at the expense of stringent ethical considerations and the protection of fundamental human rights (Castets-Renard & Besse, 2022, p.22). The economic benefits, such as job creation and technological advancement, are prioritized to maintain the EU's position in the global AI landscape, which may lead to a compromise on addressing moral and human rights issues comprehensively.

Similarly, Smuha et al. (2021) argue that the EU AI Act lacks adequate safeguards to protect fundamental rights effectively. They raise concerns about the Act's insufficient measures against AI-enabled social scoring practices, and the use of biometric identification systems. The EU AI Act prohibits AI systems from being used by public authorities for social scoring. However, social scoring is often conducted by private actors with access to vast amounts of personal data in areas such as job applications, hiring policies, and loan applications, which can have devastating effects (Smuha et al., 2021, p.23). The EU AI Act restricts remote biometric identification systems (RBIS) primarily for law enforcement in public spaces, with broad exceptions, but does not extend these prohibitions to other public or private actors (Smuha et al., 2021, p.25). While public law enforcement agencies face stringent regulations, private entities are only categorized as high-risk, creating an asymmetrical protection framework (Smuha et al., 2021, p.25). Additionally, biometric categorisation systems (BCS) and emotion recognition systems (ERS) are not adequately regulated as BCS and ERS are only subjected to transparency obligations (Smuha et al., 2021, p.26).

In line with these criticisms, Ebers et al. (2021) propose more stringent measures to address ethical concerns related to AI technologies. They advocate for banning manipulative AI systems like deepfakes and extending the ban on social scoring to private entities (p.601). Furthermore, they call for a general prohibition on biometric identification and categorization systems, including their private use, as well as the prohibition of emotion recognition systems (Ebers et al., 2021, p.601).

Shaelou & Razmetaeva (2024) highlight concerns about the potential intrusion into individuals' privacy through the collection and analysis of biometric data, particularly by public authorities (Shaelou & Razmetaeva, 2024). The utilization of AI technologies by public authorities in public spaces has the potential to exceed the boundaries of what is deemed acceptable in a democratic society that upholds the principles of the Rule of Law and European values, as well as fundamental human rights (Shaelou & Razmetaeva, 2024, p.6). AI-driven surveillance systems expose the public to heightened risks of power imbalances, granting public authorities excessive access to sensitive information about individuals' private lives (Shaelou & Razmetaeva, 2024, p.9). This not only allows for an overly detailed understanding of an individual's behaviors and habits but also creates the potential for highly targeted interventions if such data is utilized beyond the stated objectives of public authorities. AI systems have the capability to influence the information individuals encounter online, whether through search engines, news media, websites, or social media platforms (Shaelou & Razmetaeva, 2024, p.9). This control over digital content presents opportunities for manipulation by dominant players in the digital sphere, particularly large tech corporations. The potential impact of AI systems, guided by algorithms developed and maintained by these companies, extends beyond mere data control discussed during the early days of GDPR. Their ability to shape public opinion, predict behavior, and operate in a seemingly impersonal manner heightens concerns regarding responsibility and accountability (Shaelou & Razmetaeva, 2024, p.10).

Veale and Borgesius (2021) identify several enforcement and implementation challenges within the Draft EU AI Act. The Act mandates that users of these systems disclose their use to individuals, except for crime prevention purposes. This requirement is criticized as redundant, given existing data protection laws, and it may lack clarity. Additionally, it could apply to systems that claim not to process personal data, raising contentious issues (Veale & Borgesius, 2021, p. 107). Prioritizing transparency here fails to address broader concerns

regarding the scientific accuracy and societal implications of these technologies (Veale & Borgesius, 2021, p. 107). Secondly, The Act introduces a rule requiring users to disclose when content (images, audio, or video) has been altered to appear genuine, with certain exceptions. This provision aims to mitigate the harm caused by misleading AI-generated content, but its enforcement is problematic. Users may not always recognize the artificial nature of content, making enforcement complex (Veale & Borgesius, 2021, p. 107). Investigating undisclosed deepfakes requires specialized expertise that current authorities often lack (Veale & Borgesius, 2021, p. 107). Moreover, the rule's broad scope might necessitate disclosures for benign activities like AI-generated stock photos, raising practical concerns (Veale & Borgesius, 2021, p. 107). There is also potential overlap with existing laws addressing issues like fraud (Veale & Borgesius, 2021, p. 107). Lastly, The Draft AI Act covers a wide range of AI technologies but imposes extensive obligations on only a subset considered high-risk. This creates a disparity where the Act's broad scope impacts numerous AI systems, yet its stringent requirements target only a few. This approach, combined with the Act's complexity and narrow focus, may lead to enforcement and compliance challenges, potentially hindering the Act's overall effectiveness in achieving its regulatory goals within the EU (Veale & Borgesius, 2021, p. 107).

Laux, Wachter, and Mittelstadt (2024) criticize the EU AI Act for conflating trustworthiness with risk acceptability, arguing that this oversimplification undermines evidence-based policymaking. They advocate for a more participatory approach to public accountability and highlight the challenges posed by citizens' limited technical knowledge and the domain-specific nature of trust in AI (Laux et al., 2024, p. 27-28).

Various organizations, including the European Centre for Not-for-Profit Law (ECNL), Liberties, the European Civic Forum (ECF), Access Now, AlgoRights, Amnesty International, Article 19, EDRi, the Irish Council for Civil Liberties, the European Disability Forum, AlgorithmWatch, Bits of Freedom, Fair Trials, LEFÖD, Kermes, and Politiscope, have critically assessed the EU AI Act. They argue that the Act prioritizes industry, security services, and law enforcement interests over the protection of the rule of law and civic space.

One major concern is the excessive discretion allowed by the Act to the European Commission, secondary legislation, and voluntary codes. This flexibility can undermine safeguards and erode fundamental rights over time (ECNL, Liberties, & ECF, 2024, p.5).

Specifically, the Act's prohibitions on real-time remote biometric identification and biometric categorization contain broad exceptions for law enforcement, potentially rendering these bans ineffective. This could lead to abuses such as surveillance of peaceful assemblies and harassment of protesters (ECNL, Liberties, & ECF, 2024, p.5-6).

Another issue is the self-assessment requirement for high-risk AI system providers. This allows companies and authorities to unilaterally decide that their systems pose no significant risks, bypassing oversight and potentially leading to discriminatory practices in public benefits or law enforcement (ECNL, Liberties, & ECF, 2024, p.6).

The Act mandates listing potential impacts on fundamental rights but lacks a clear obligation to prevent or mitigate these impacts. The removal of external stakeholder consultation, including civil society input, limits transparency and accountability (ECNL, Liberties, & ECF, 2024, p.7). Additionally, the Act exempts AI systems used for national security from scrutiny, allowing governments to bypass safeguards, leading to unregulated surveillance and other rights violations, particularly in countries with compromised civic space and rule of law (ECNL, Liberties, & ECF, 2024, p.8).

Civic participation in the Act’s implementation and enforcement is limited. Civil society can represent individuals only in specific consumer rights cases, and meaningful engagement in fundamental rights impact assessments is not guaranteed. The advisory forum within the AI Office and AI Board is the only formal mechanism for civil society involvement (ECNL, Liberties, & ECF, 2024, p.9).

Other critical points include the AI Act’s inadequate coverage of accessibility requirements, which should apply to all risk levels of AI systems to better serve people with disabilities (Joint Statement, 2021, p.3). The public EU database for high-risk AI systems has significant transparency limitations, as private sector deployers and certain public authorities are not required to provide comprehensive information, hindering public scrutiny and accountability (Joint Statement, 2021, p.3).

While the inclusion of Fundamental Rights Impact Assessments (FRIAs) is a step forward, the Act lacks mandates for meaningful assessment and prevention of negative impacts, does not require engagement with external stakeholders, and provides transparency exceptions for

law enforcement and migration authorities (Joint Statement, 2021, p.4). The "remedies" chapter is also criticized for its insufficient enforcement mechanisms and limited access to meaningful explanations from AI deployers, undermining the efficacy of remedies for affected persons (Joint Statement, 2021, p.4-5).

The broad exemptions for AI systems used for national security purposes bypass crucial human rights safeguards, posing significant risks to fundamental rights through unchecked use of biometric surveillance (Joint Statement, p.6). Although the Act partially bans biometric and emotion recognition systems, significant exceptions for law enforcement and migration authorities could legitimize intrusive surveillance practices and AI-fueled discrimination, particularly against marginalized groups (Joint Statement, 2021, p.6).

Finally, the Act's failure to ban retrospective facial recognition and its insufficient safeguards for such uses allow continued rights violations under lower thresholds than current EU data protection laws (Joint Statement, 2021, p.7).

Chapter II:

AI Policies and Practices in Russia

2.1 Introduction

In this chapter the way how the use of artificial intelligence (AI) is legislatively regulated in Russia will be explored and how these technologies are currently utilized by the state to ensure national security, which is prioritized over citizens' freedoms. In the first part of this chapter, we will delve into the legal framework governing AI in Russia. To ensure accuracy, a semi-structured interview was conducted with a Russian lawyer, Interviewee 1 (here and after the name was changed in order to protect the respondent), who provided detailed insights into the legal aspects of Russian legislation in this area. The second part will examine the emergence of new systems that can automatically identify violators based on data from the Internet and how the AI systems are used for protection.

Despite advancements in legislative and regulatory acts regarding artificial intelligence (AI) use in Russia, significant gaps in legal regulation persist, undermining effective protection of citizens' rights and interests. While experimental legal regimes, mandatory insurance for AI-induced damage, and the development of ethical guidelines represent positive steps toward a legal framework for AI, deficiencies remain. There's a lack of a comprehensive approach to AI-related issues and insufficient attention to compensation and guarantee concerns for citizens.

Moreover, while legislation mandates AI's respect for human freedom and privacy, the current regulatory framework and infrastructures are heavily influenced by Russia's political situation and international tensions due to the war in Ukraine. State initiatives like "Clean Internet" exploit legal loopholes under the guise of national security. Systems like Oculus and Vepr, ostensibly for national safety, are instead utilized for surveillance, monitoring internet users, and suppressing regime dissent, even employing pro-government chatbots. Biometric identification, omnipresent across various services, including medical and banking sectors, facilitates police surveillance, extending to the identification and apprehension of peaceful protesters.

In response to challenges like deep fakes, the government has intensified efforts to combat anonymity, expanding surveillance measures. These actions, while purportedly aimed at enhancing safety and addressing deepfake concerns, incrementally erode privacy and civil liberties. The use of AI for controlling content, influencing public opinion, and combating political opposition reflects a broader strategy of utilizing AI for State control, necessitating a careful balance between security imperatives and individual freedoms.

The OpenAI report highlights the use of AI in Russian influence operations, notably through networks like Bad Grammar and Doppelganger. These campaigns target audiences

internationally, using AI-generated content to spread political narratives across platforms like Telegram and social media. Bad Grammar focuses on generating political comments in Russian and English, while Doppelganger spreads anti-Ukraine content through various websites and social media channels. Despite their efforts, these campaigns often struggle to gain significant engagement, indicating limitations in their effectiveness.

2.2 Legal regulation of AI in Russia

The State interest in the development of AI technologies, and the need for its legal regulation, arose, according to the lawyer Interviewee 1, in 2019. During this year the President of the Russian Federation issued Decree No. 490 of October 10, 2019, "On the Development of Artificial Intelligence in the Russian Federation," which approved the National Strategy for the Development of Artificial Intelligence in the Russian Federation until 2030.

In Russia, there is a national strategy for the development of artificial intelligence until 2030 and an alliance group involved in AI work, united to promote the integration and development of AI in Russia. The alliance brings together major Russian companies, including state-owned ones, to jointly develop competencies and accelerate the implementation of artificial intelligence in education, scientific research, and business practices. The Alliance unites major Russian companies, including state-owned enterprises, for the joint development of competencies and accelerated implementation of artificial intelligence in education, scientific research, and business practices. The Alliance includes companies such as:

- Uralchшm: a leading producer of mineral fertilizers.
- Sberbank: the largest bank in Russia and Eastern Europe.
- Severstal: one of the world's largest steel producers.
- Gazprom: the largest company in Russia and one of the largest in the world in the extraction and sale of natural gas.
- Yandex: a leading technology company providing internet services and products.
- Russian Direct Investment Fund: a sovereign fund that engages in direct investments in Russian companies.
- VKontakte: the largest social network in Russia.
- Sibur: a leading company in the petrochemical industry.

The alliance aims to become the center for the development of artificial intelligence in Russia, ensuring the country's technological leadership and the participating companies' global market competitiveness. Its main areas of activity include education, regulatory improvement, communication, investment attraction, and AI application.

The alliance has created and actively developed the AI Russia project, aimed at demonstrating operational Russian cases using artificial intelligence.

As for the National Strategy for the Development of AI in Russia, the main principles of AI development and integration into society include the protection of human rights and freedoms; security (preventing the deliberate use of AI to harm citizens and organizations, preventing and minimizing the risks of negative consequences of AI technology use, such as confidentiality breaches and disclosure of restricted information), as well as using AI to ensure information security (p.8); transparency; technological sovereignty (Russia's independence in AI through the independent development and use of domestic technologies); promoting close cooperation between scientific research and AI development; most effective AI technology use (prioritizing existing mechanisms for implementing state policy in scientific and technical areas and other fields); supporting competition and preventing monopolization (Decree of President №490, 2019, p.9).

The main objectives for the development of artificial intelligence in the Russian Federation (Decree of President №490, 2019, p.10) are:

1. Increasing the availability of infrastructure for the development of AI technologies.
2. Supporting organizations developing AI technologies.
3. Promoting scientific research and development for the advanced growth of AI.
4. Enhancing competencies and raising public awareness about AI technologies.
5. Encouraging the integration of AI technologies into the economy and social sphere.
6. Mandating the implementation of trusted AI technologies in areas where their use could impact the security of the Russian Federation.
7. Creating a regulatory framework to ensure the safe application of AI technologies.
8. Strengthening international cooperation in the field of AI.

According to the commentary of the lawyer, Interviewee 1, to date, there is no unified (codified) federal law in Russian legislation dedicated to regulating artificial intelligence. However, individual provisions can be found in related branches of legislation.

For example, in 2020, the Federal Law N 123-FZ "On Conducting an Experiment to Establish Special Regulation to Create the Necessary Conditions for the Development and Implementation of Artificial Intelligence Technologies in the Subject of the Russian Federation - the Federal City of Moscow and Amending Articles 6 and 10 of the Federal Law" was adopted.

Point 2 of Article 2 of the aforementioned Federal Law provides a definition of "artificial intelligence." Thus, artificial intelligence is a set of technological solutions that allow for the imitation of human cognitive functions (including self-learning and problem-solving without a pre-defined algorithm) and obtain results comparable, at a minimum, to the results of human intellectual activity. The set of technological solutions includes information and communication infrastructure (including information systems, information and telecommunications networks, other technical means of information processing), software (including those using machine learning methods), data processing processes and decision-making services.

Thus, the application of artificial intelligence (AI) is partly regulated by the Federal Law "On Personal Data," which establishes the following.

According to subparagraph 9.1 of paragraph 1 of Article 6 of the Federal Law "On Personal Data," the processing of personal data must be carried out in compliance with the principles and rules provided by this Federal Law. Processing of personal data is allowed in the following cases: processing of personal data obtained as a result of depersonalization of personal data is carried out to increase the efficiency of state or municipal management, as well as for other purposes provided by Federal Law N 123-FZ of April 24, 2020, "On conducting an experiment to establish special regulation to create the necessary conditions for the development and implementation of artificial intelligence technologies in the subject of the Russian Federation - the federal city of Moscow and amending Articles 6 and 10 of the Federal Law 'On Personal Data' and Federal Law N 258-FZ of July 31, 2020, "On

experimental legal regimes in the field of digital innovations in the Russian Federation," in the manner and on the terms provided by the specified federal laws.

In accordance with paragraph 2.1 of Article 10 of the Federal Law "On Personal Data," the processing of personal data relating to health status obtained as a result of depersonalization of personal data is allowed to increase the efficiency of state or municipal management, as well as for other purposes provided by Federal Law N 123-FZ of April 24, 2020, "On conducting an experiment to establish special regulation to create the necessary conditions for the development and implementation of artificial intelligence technologies in the subject of the Russian Federation - the federal city of Moscow and amending Articles 6 and 10 of the Federal Law 'On Personal Data' and Federal Law N 258-FZ of July 31, 2020, "On experimental legal regimes in the field of digital innovations in the Russian Federation," in the manner and on the terms provided by the specified federal laws.

It should also be noted that the Government of the Russian Federation, by its decree of August 19, 2020, No. 2129-r, approved the Concept for the development of regulation of relations in the field of artificial intelligence and robotics until 2024 (hereinafter - the Concept), developed by the Ministry of Economic Development of Russia.

According to paragraph 1, the purpose of the Concept for the development of regulation of relations in the field of artificial intelligence and robotics is to determine the basic approaches to transforming the system of regulatory regulation in the Russian Federation to create and apply such technologies in various sectors of the economy while ensuring the rights of citizens and ensuring the security of individuals, society, and the state (Decree of Govt. N 2129-r, 2020). The Concept also aims to establish an environment conducive to laying the groundwork for legal regulation concerning emerging social dynamics linked to the advancement and utilization of artificial intelligence and robotics technologies and their derivative systems. Additionally, it seeks to pinpoint legal obstacles impeding the advancement and utilization of such systems (Decree of Govt. N 2129-r, 2020).

The Concept seeks to foster the growth, integration, and utilization of these technologies, fostering the creation of artificial intelligence and robotics systems that operate in secure and reliable capacities. This endeavor aims to drive robust economic expansion, enhance the well-being and living standards of individuals, safeguard national security and public safety,

and secure enduring competitiveness for the Russian economy, positioning it prominently on the global stage within the realm of artificial intelligence (Decree of Govt. N 2129-r, 2020) .

In his interview, Interviewee 1 noted that one of the interesting directions of the Concept is the elaboration of issues related to responsibility for various offenses caused by the use of artificial intelligence technologies. Moreover, the Concept indicates not only civil liability but also criminal and administrative liability, which fully meets the requirements of the Constitution of Russia.

The Concept became the first regulatory legal act (subordinate legal act) aimed at the legal development of the field of artificial intelligence.

The development of artificial intelligence in Russia is also supported through the introduction of experimental legal regimes (hereinafter referred to as ELRs), which have their legal basis:

- Federal Law N 258-FZ of July 31, 2020, "On experimental legal regimes in the field of digital innovations in the Russian Federation";
- Federal Law N 331-FZ of July 2, 2021, "On Amending Certain Legislative Acts of the Russian Federation in Connection with the Adoption of the Federal Law 'On Experimental Legal Regimes (ELR) in the Field of Digital Innovations in the Russian Federation'."

As of today, the Government of the Russian Federation has adopted 13 resolutions dedicated to the application of artificial intelligence in the fields of unmanned transport and healthcare.

The objective of these ELRs, as outlined in Section IV, encompasses fostering competition; broadening the range, improving the quality, or increasing the accessibility of products, services, and endeavors; fostering advancements in science and societal domains; refining overall regulatory frameworks based on the outcomes derived from ELR implementation; and establishing conducive environments for the emergence and adoption of digital innovations.

The document acknowledges (paragraph 40) that during the implementation of these ELRs, risks of personal data leakage of patients and limited access information may arise due to cyberattacks, technical failures, violation of personal data processing rules, information

protection and information security regime, rules for operating information platforms, devices, medical information systems, support systems, dishonest actions of subjects, and participants of ELRs.

Additionally, the document addresses risks of harm to patients' lives and health from the implementation of ELRs, such as prescribing incorrect treatment by attending physicians due to the low diagnostic value of data obtained from the device due to incorrect device operation.

Regarding violations of personal data processing (e.g., leaks), the document refers to the norms provided by the Federal Law "On Personal Data," which since 2020 also includes the collection and processing of personal data using artificial intelligence technologies.

According to the opinion of the Interviewee 1, individuals affected by such ELR programs simply do not understand the legislative framework regulating artificial intelligence technologies due to the lack of specialized federal legislation. Without qualified legal support, citizens will simply not understand when their rights are violated, for example, in the processing of personal data using artificial intelligence, or when improper services are provided in the field of medicine and healthcare. From a legal perspective, legislation and legal regulation should be accessible to all individuals without exception.

On February 15, 2024, the State Duma of the Russian Federation adopted in the first reading a bill on compulsory insurance for damage caused by artificial intelligence within the framework of ELRs. Currently, risk insurance within ELRs is voluntary. The State Duma proposes to obligate ELR subjects to insure their liability. According to the Interviewee 1, this will increase the level of protection for individuals affected by ELRs, as compulsory insurance of ELR subjects will allow individuals harmed by the implementation of the program to recover losses through the insurance organization.

Another innovation is the creation of a commission to identify the circumstances under which harm is caused by the application of artificial intelligence. It is proposed that the composition of this commission will include representatives of authorized and regulatory bodies, business community organizations, and other individuals. The procedure for the formation and operation of such a commission will be determined by a regulatory legal act of the authorized

body, and at the direction provided for in paragraph 4 of part 2 of Article 1 of the Federal Law "On Experimental Legal Regimes in the Field of Digital Innovations in the Russian Federation" - by the Bank of Russia.

According to legal experts, the adoption of the law will improve the level of legal regulation of the use of artificial intelligence technologies.

In 2023, the Central Bank of the Russian Federation issued its recommendations on "Application of Artificial Intelligence in the Financial Market." As stated in the recommendation, the financial sector is among the sectors of the economy where AI technologies are being actively implemented. The Central Bank notes that artificial intelligence technologies are used in various directions: scoring, underwriting, trading, investment consulting, risk management, fraud prevention, and more.

Regarding the regulation of artificial intelligence technologies, the Bank of Russia proposes a proportionate regulation, based on a risk-oriented approach. This approach implies adjusting the format of regulatory requirements based on the volume and probability of the risk of using the technology compared to the potential positive effect of its implementation. The use of a risk-oriented approach will allow maintaining a balance between creating conditions for the development of AI and risk management, requiring attention from the regulator. In other words, the Bank of Russia differentiates risks from the use of artificial intelligence into those requiring regulatory intervention and those that do not.

In addition, Russia has adopted the Code of Ethics in the Field of Artificial Intelligence, which is advisory in nature but nonetheless contributes to the regulation of artificial intelligence (AI) technology use. The objectives of the Code are:

- To provide recommendations for making ethical decisions regarding the creation and use of artificial intelligence.
- To reduce the risks of unethical use of artificial intelligence that violate human rights and interests.
- To create a tool for interaction between the government, developers, scientific organizations, and society on issues of AI ethics.

From Interviewee 1’s perspective, among the current shortcomings is the lack of comprehensive legal regulation of AI technology use, including the following issues:

1. Compensation for harm to life and health from AI use.
2. Compensation for material damage from AI use.
3. The process of proving harm caused by AI use.
4. Issues of holding entities accountable under civil, administrative, and criminal law.
5. Guarantees for individuals affected by AI use, including the protection of rights and lawful interests of workers, which, in the author's view, should not be subject to sudden reduction due to AI technology advancements.
6. Introduction of anti-abuse norms in the financial sector.

Interviewee 1 states: "*Federal legislation in this area is evolving, as evidenced by the adoption of a number of recommendatory acts by state bodies, as well as the implementation of numerous ELR regimes in the territory of the Russian Federation and its subjects. Moreover, the interest in developing legal regulation is confirmed by the bills considered by the State Duma, including those on compulsory insurance of ELR subjects' liability. As for further regulation, Russia needs to adopt a special Federal Law dedicated to artificial intelligence technologies*."

In analyzing legislative and regulatory acts related to the use of artificial intelligence (AI) in Russia, significant steps towards regulating this area can be noted. The introduction of experimental legal regimes, mandatory insurance for AI-caused damage, as well as the development of recommendations and codes of ethics, are important steps in forming a legal framework for the development and use of AI technologies. However, despite the measures taken, significant gaps remain in legal regulation, such as the absence of a comprehensive approach to problems related to AI use, insufficient attention to compensation and guarantee issues for citizens. Further improvement of legislation and the development of specialized regulatory acts are necessary to ensure effective protection of the rights and interests of all participants in the process of applying artificial intelligence in various spheres of public life.

2.3 *AI systems safeguarding national security*

In this section I will delve into the emergence of new systems that utilize artificial intelligence to automatically identify violators based on data from the Internet. We will explore how the Russian government employs these AI technologies to enhance state security. This includes an examination of advanced monitoring systems like "Oculus" and "Vepr", which analyze online content and user behavior, as well as the use of biometric data for law enforcement purposes. These systems highlight the intricate balance between leveraging AI for public safety and the implications for individual privacy and freedoms.

To provide the reader with a comprehensive understanding before delving into the details, it is important to give an overview of the two main organizations that manage the digital sphere in Russia. To combat harmful content on the internet in Russia, there is a federal executive body responsible for regulating communications, information technology, and mass media - Roskomnadzor. Roskomnadzor is involved in the development and implementation of state policy and also regulates the sphere of information technology, postal and electronic communications, mass media, printing, and processing of personal data. It is also responsible for monitoring compliance with copyright laws and blocking content that may harm health and children's development. The GRChTs was established in 2000 based on the disbanded State Communications Supervision Service to control the use of radio frequencies by telecommunication operators. In 2008, the center came under the management of Roskomnadzor. In May 2014, the GRChTs was tasked with monitoring compliance with legislation in the sphere of Roskomnadzor's activities.

Information about the development of the mentioned above systems became known in November 2022 when the Belarusian group "Cyber Partisans" hacked one of the structures of Roskomnadzor - the Main Radio Frequency Center (GRChTs) and disclosed an archive of documents and emails downloaded from the center. Documents disclosed by hackers prove the existence of systems engaged in cyber espionage and compiling reports on citizens. Roskomnadzor confirmed the hack but stated that the intruders did not manage to gain access to critical infrastructure (Radio Liberty, 2022). The presence of systems disclosed by hackers can also be found in documents published by the state: on the government procurement website, government orders, and technical tasks and appointments of these systems.

At the end of 2019, the Federal State Unitary Enterprise "Main Radio Frequency Center" (GRChTs), responsible for monitoring mass media, mass communications, and information technologies in accordance with Russian legislation, developed a messenger called "Cabinet for Operational Interaction" (Federal State Unitary Enterprise, 2019). This messenger is intended for secure and convenient communication between employees of Roskomnadzor and GRChTs with representatives of law enforcement agencies such as the General Prosecutor's Office, the Federal Security Service (FSB), the Federal Protective Service (FSO), the National Guard, and the Ministry of Internal Affairs (MVD).

In the group chats of this messenger, GRChTs specialists publish reports on "protest sentiments" and monitor instances of destabilization in Russian society, while in private messages, they send links to posts by Russians containing "fakes" about the Russian army to employees of the General Prosecutor's Office. Since the beginning of the conflict, GRChTs, in collaboration with Roskomnadzor, has identified over 160,000 such "fakes" (BBC News, 2023). Reports from the organization indicate that in 2020, as part of the program "From a Unified Information Space to a Unified Departmental Digital Platform," Roskomnadzor planned to allocate an additional 7 million rubles for the creation of a mobile application for the messenger "Cabinet for Operational Interaction" (Federal State Unitary Enterprise "Main Radio Frequency Center," 2020). However, the question remains: how did Roskomnadzor and GRChTs increase their productivity in searching for "unlawful content"?

A representative of Roskomnadzor commented to the news outlet "Vedomosti" that previously, all images, websites, and videos were manually processed by operators at the Main Radio Frequency Center, a subdivision of Roskomnadzor. On average, around 106 images and 101 videos were processed per day. However, with the introduction of the "Oculus" system, the number of processed files skyrocketed to over 200,000 images per day (Tunyayeva M., 2023).

"Oculus" is a system designed for classifying images and videos according to specified requirements, including basic types of prohibited content (Tunyayeva M., 2023). "Oculus" functions as a classifier working with pre-defined sources of information, where content is analyzed for compliance with legislative norms (Tunyayeva M., 2023). In other words, "Oculus" analyzes photo and video materials collected by other services that gather data from

specific web pages, publics, or social media profiles (Tunyayeva M., 2023). The program does not collect data independently but only classifies it. The system is developed to detect and block propaganda of drug use, calls for suicide, child pornography, as well as fakes - materials aimed at substituting real facts with specially constructed "reality" (Tunyayeva M., 2023). In 2022, upon the request of the Prosecutor General's Office of Russia, over 102,000 internet resources containing unreliable information, including information about the course of the Special Military Operation in Ukraine, were removed and blocked, representatives of the Roskomnadzor press service told "Vedomosti" (Tunyayeva M., 2023). For comparison, in 2021, there were 7,203 such resources, in 2020 - 1,525, and in 2019 - 311 (Tunyayeva M., 2023).

In addition to drug propaganda and child pornography, "Oculus" also blocks LGBT propaganda, extremist themes, and calls for mass illegal activities. Moreover, according to the technical specifications, in addition to digital content, "Oculus" analyzes chats and channels in messengers, inscriptions, URL addresses, subtitles, and QR codes. Analogues of such systems continue to be developed, including, for example, "Demon Laplace" and "Monitoring of Depressive and Suicidal Behavior in Children and Adolescents" (RosComFreedom 2018, 2021).
Such programs can track comments, personal messages, and private channels if internet services transmit data to Roskomnadzor. There is a registry of information distribution organizers who are required to provide data upon the request of the FSB. This registry includes both Russian and foreign services such as VKontakte, Yandex (Mail, Metrica, Taxi, and others), Odnoklassniki, Tinder, YouTube, and Telegram. According to Federal Law No. 97-FZ dated May 5, 2014, these services are obliged to collect, store, and provide information, including not only metadata but also IP addresses, geolocation, internal messages, files, and contacts. The operation of this system should be regulated by the Personal Data Law N 152, Article 10, which prohibits processing data containing opinions related to racial and national origin, politics, religious or philosophical beliefs, health status, and personal life. However, in the case of a threat to Russia's security, this information can be processed without the user's consent, and clarifications on this matter can only be provided by the authorized body - Roskomnadzor.

Can we rely on the objectivity of the determinations made by Roskomnadzor? To answer this question, let's look at Roskomnadzor's recent achievements in this area. According to the

article on army fakes (Article 207.3), all materials contradicting the official position of the Ministry of Defense and the Russian government are considered criminal lies. Moreover, the "fakes" themselves are classified into various categories, including "killing civilians," "shelling civilian populations," "losses of personnel and equipment," and others. Each year, punishments under this article become increasingly severe. For example, while violators only received fines in 2022, by 2023, imprisonment became possible. According to Forbes data, in the first six months of 2023, Russian courts sentenced 21 people for army fakes, eight of whom were sentenced to actual imprisonment, while another three received suspended sentences, and ten were fined (Baturov, T., 2023). In addition, there are articles on discrediting the army (Article 280.3) and on calls for sanctions against Russia (Article 284.2 of the Criminal Code of the Russian Federation), which also entail punishments, including imprisonment. By the end of 2022, 4,440 people had been convicted for violations of these articles (Kornya, A., 2023).

One of the most high-profile cases related to this article was the case of Sasha Skolichenko. The St. Petersburg artist brought price tags with inscriptions such as "Stop the war! 4,300 Russian soldiers died in the first three days" and "Putin has been lying to us on TV screens for 20 years" to the "Perekrestok" store on Vasilievsky Island. For this act of peaceful protest, she was sentenced to 7 years of imprisonment (Turkova K., 2023).

The second law widely used by Roskomnadzor is the law on foreign agents (327-FZ). Over the past three years, the number of foreign agents in Russia has increased by 3.5 times (Sokolova, M., 2024). They are prohibited from working in the police, teaching in schools and universities, as well as receiving state funding. Their information products must be labeled, and citizens face fines for cooperating with them. Previously, foreign agents were identified based on criteria such as funding from abroad and opinions expressed by them, which often contradicted the views of official government representatives. However, since 2022, funding from abroad has ceased to be a key factor, and at present, foreign agents are chosen situationally (Verstka, 2023).

The article on foreign agents imposes liability only in case of repeated violations or if a person has previously been held accountable for similar articles (Verstka, 2023). Foreign agents face persecution: in 2022, 38 out of 209 foreign agents were under persecution, and in 2023, out of 406 foreign agents, 86 are being persecuted (Verstka, 2023).

In today's conditions, any expression of dissatisfaction with state policy can easily lead to serious consequences. For example, after Konstantin Petrov published a video criticizing Russian cities, claiming that there is nothing to see in them, he was accused of rehabilitating Nazism based on a report from the Safe Internet League. As a result, he faces up to 5 years of imprisonment (Turkova K., 2023). This confirms that there are no guarantees of fair determination of harmful content.

In conjunction with the Oculus system, the Vepr system operates. The information system "Vepr" is designed to identify potential points of tension on the Internet that may lead to information threats, and to analyze and predict the further spread of destructive materials. According to a statement from the Roskomnadzor press service (Gavriluk A., & Rozhkov R. 2023), it is envisaged that "Vepr" will collect textual, audio, video, and graphical information from resources with a daily audience of at least 1 million people, including online platforms and forums. The system will analyze no fewer than 20,000 media materials per day and publications from no fewer than 300 media outlets within 20 minutes of their appearance (Gavriluk A., & Rozhkov R. 2023). All incidents found will be systematized into cards with links, publication dates and times, hashtags, authors, platform owners and hosting, as well as the number of views and likes. "Vepr" will generate analytical reports, search for incidents based on specified criteria, forecast the spread of fakes, identify dangerous media campaigns, and detect and stop bot networks.

The Clean Internet initiative, introduced by Roskomnadzor and the Main Radio Frequency Center (GRChTs), appears to be more than just a PR campaign to combat anti-government content. In mid-2022, leaked documents revealed a plan and project "passport" officially titled Clean Internet. According to these documents, the project involves the creation of what could be described as a "bot farm" (Agentstvo, 2023). This bot farm is described as a "software-hardware complex for automated creation and management of accounts on social networks," and a bot account is defined as a "program that performs actions automatically and/or on a schedule, having certain similarities to a human" (Agentstvo, 2023).

Roskomnadzor and GRChTs explain that the bot farm is necessary to reduce the time spent creating accounts on social networks, expedite the posting of posts and comments, and centralize all accounts in one place (SVTV, 2023). One of the main tasks of the "Clean

Internet" system is to populate the created accounts with content to simulate user activity, including photos, posts, and personal data. In simpler terms, this is a state-run system for creating deepfakes. The bot farm is designed to operate in a way that the bots successfully pass verification when joining closed groups and communities. The system will generate unique IP addresses for each bot account and impose restrictions on the number of friends, reposts, likes, and group memberships. According to Roskomnadzor documents, the bots are expected to behave in a manner that their lifespan in public communities is no less than three months and in closed ones no less than a month (SVTV, 2023). However, how exactly will the bots determine how to accurately mimic user behavior to successfully masquerade as humans? For this purpose, the Vepr system is utilized—it evaluates the audience size and the likelihood of discussion growth. Then, upon detecting heated discussions, the bot farm starts shaping the "correct public opinion." Thus, the joint operation of these systems ensures comprehensive control over public opinion.

Oculus interacts with Vepr and another Roskomnadzor system called "Mir" (World). Mir collects messages from various sources and classifies them by content types, including texts, videos, and images. Oculus analyzes images and videos to search for materials that violate legislation, while Vepr analyzes texts and predicts the further dissemination of the topic. Although automatic submission of materials to law enforcement agencies for initiating cases and their inclusion in the registry of prohibited sites will not be carried out, access to the data through APIs will be provided to third parties, including law enforcement officials (SVTV, 2023). This will enable the identification of more violators, although the punishment process will remain non-automated.

As for the international tools to spread propaganda by AI and promote Russian national interests, Bad Grammar and Doppelganger campaigns are important to mention. The OpenAI report from May 2024 examines the latest trends in AI-driven influence operations. On the Russian side such networks as Bad Grammar and Doppelganger were described. A previously unreported Russian network, dubbed "Bad Grammar," targeted audiences in Russia, Ukraine, the United States, Moldova, and the Baltic States. This network used AI models to generate political comments in both Russian and English, which were then posted on Telegram. The network struggled to build an audience and often posted ungrammatical English comments (OpenAI, 2024, p.13). Its activities included setting up automated comment-spamming pipelines and generating content from multiple fake personas, particularly focusing on pro-Russia Telegram channels (OpenAI, 2024, p.14). The primary

topics addressed by this network included the war in Ukraine, the political situation in Moldova and the Baltic States, and U.S. politics (OpenAI, 2024, p.15). Russian-language comments frequently accused the presidents of Ukraine and Moldova of corruption and betrayal, while English-language comments focused on immigration, economic hardship, and current events, often arguing against U.S. support for Ukraine (OpenAI, 2024, p.16).

The "Doppelganger" influence operation involved multiple clusters of accounts generating anti-Ukraine content across various websites and social media platforms. This campaign targeted audiences in Europe and North America and utilized different tactics and techniques across clusters. The first cluster focused on short text comments in several languages, while others translated and published articles on attributed websites such as rrn[.]media and franceeteu[.]today (OpenAI, 2024, p.21). The campaign also attempted to evade detection by accessing AI models through a service linked to companies in Russia and the Czech Republic (OpenAI, 2024, p.22). The majority of Doppelganger's content portrayed Ukraine, the US, NATO, and the EU negatively, while casting Russia in a positive light (OpenAI, 2024, p.22). Common themes included the supposed economic importance of Russia to Europe, the disconnection of Western leaders from their voters, and negative portrayals of Ukraine as weak and corrupt (OpenAI, 2024, p.22).

The Ministry of Internal Affairs of Russia (MVD) plans to utilize advanced neural networks for identifying offenders. In 2024, the department will conduct relevant research, with the development of two smart technologies slated for the following year (Ustinova A., Kinyanina E, 2024).

The first of these systems, tentatively named Conjuncture, will forecast emergencies and negative incidents while simulating response scenarios. A Western counterpart to Conjuncture is IBM's Watson Openscale service, which monitors and manages AI models (Ustinova A., Kinyanina E, 2024).

The implementation of these plans will be carried out by employees of new departments to be established by the end of 2024 in two organizations under the control of the MVD. Additionally, a special data analysis laboratory will be established to serve as a "sandbox"—an isolated environment for the safe use of programs. This laboratory will handle data management and modeling. Sandboxes allow for the efficient training of neural network

models on large datasets while ensuring security and control over the process. The government already uses data from cameras integrated into the "smart city" system to train neural networks. This information, along with data from open sources and MVD's proprietary data, aids in the development and improvement of artificial intelligence technologies. The "Safe City" system comprises video surveillance complexes connected to a unified network managed by artificial intelligence.

The system's operation is based on a database containing photos of wanted individuals. When a person passes through turnstiles, cameras capture their face and instantly transmit the image to the system. Artificial intelligence compares this image with photos in the database in a fraction of a second. If a match is detected, the system notifies the relevant authorities, who can promptly take action. In the Moscow metro, the system is used experimentally, and the police already recognize its effectiveness. Previously, the MVD used cameras to identify offenders. For example, after protests in support of Navalny in 2021, protesters in Moscow were mass administratively punished for participating in protests based on video recordings from city surveillance cameras (A.Alexandrov, S. Korsakov, 2021). Also, in 2022, video surveillance systems were upgraded to identify violators of the coronavirus regime (A.Alexandrov, S. Korsakov, 2021).

However, access to biometrics is not only taken from surveillance cameras. Biometric identification is already actively implemented in various spheres of civil life. Biometric data was collected from a person's voluntary consent for foreign passports, access to government services, medical institutions, ATMs, and Yandex services. Starting from June 1, 2023, all institutions will be required to transfer this data to the Unified Biometric Center. It will simplify identity verification: verified biometric data used for facial and voice recognition may replace passports in cases where presenting them for identity confirmation is currently required.

The second system, "Clone," is designed to detect fake video images for law enforcement purposes. An international counterpart to the "Clone" system is the Deepfake Detection Challenge developed by companies such as Microsoft and Amazon. It uses artificial intelligence algorithms to detect fake videos (Ustinova A., Kinyanina E, 2024).

A bill with amendments to the Russian Criminal Code is planned to be submitted to the State Duma, providing for criminal liability for creating deepfakes using voice or image forgery. This information is reported by "Izvestia" with reference to the relevant document (Bashlykova, N., & Krylova, E., 2024).

The explanatory note to the bill states that the development of computer technologies has significantly increased the capabilities of creating video and audio materials based on image and voice samples of a person (Bashlykova, N., & Krylova, E., 2024). Depending on the article, creating deepfakes may result in a fine of up to 1.5 million rubles or imprisonment for up to seven years (Bashlykova N., & Krylova E., 2024). According to officials, to address the issue of deepfakes, the primary objective lies in combating anonymity on the internet. One proposed solution involves leveraging artificial intelligence to counteract phone scams, suggesting a potential necessity for AI to monitor conversations, despite the discomfort it may entail.

Chapter III:

AI Policies and Practices in China

3.1 Introduction

This chapter synthesizes that China's approach to regulating artificial intelligence (AI) reflects a multifaceted strategy aimed at balancing technological advancement, governance objectives, and societal impact. The country has strategically positioned itself through initiatives like "Made in China 2025" and the "New Generation Artificial Intelligence Development Plan" to lead global AI innovation while addressing regulatory challenges. This includes enacting comprehensive laws such as the Data Security Law and Personal Information Protection Law, alongside regulations governing facial recognition and algorithmic recommendations, underscoring efforts to safeguard data privacy and promote transparency.

China's regulatory framework emphasizes not only technological progress but also ethical considerations and societal welfare. Despite criticisms regarding legislative fragmentation

and privacy concerns from extensive surveillance, these regulations illustrate China's commitment to managing AI's impact responsibly.

Internationally, China leverages AI for strategic purposes, including cyber-espionage and influence operations through entities like Storm-1376, showcasing AI's role in hybrid warfare and raising global security and democratic integrity concerns.

Moreover, domestically, China deploys AI in social surveillance systems like the Integrated Joint Operations Platform (IJOP), exemplified in Xinjiang, which facilitates targeted repression under the guise of counterterrorism. This highlights AI's role in reinforcing authoritarian governance practices and human rights violations.

Looking forward, China's ongoing refinement of AI ethics guidelines and international cooperation initiatives demonstrates a proactive stance in addressing ethical challenges and fostering responsible AI development. However, debates persist regarding the balance between state control and individual freedoms in China's regulatory approach, reflecting the complexities of AI governance within an authoritarian political context.

3.2 The Case of China : The Legal Framework

The government's interest towards the emerging technologies was marked with the "Made in China 2025" plan- a strategic initiative by the Chinese government aimed at transforming the nation into a global leader in high-tech industries. Issued in 2015, the primary objective of this plan is to move China up the value chain in manufacturing and reduce its dependency on foreign technology by fostering innovation and promoting advanced industries (Wübbeke et all., 2016). The emphasis of the plan is on key sectors that are critical for China's economic future, including information technology, robotics, aerospace, marine engineering, advanced rail equipment, energy-saving and new energy vehicles, power equipment, agricultural machinery, new materials, biopharmaceuticals, and high-tech medical devices (Wübbeke et all., 2016, p.6). By focusing on these industries, China seeks to enhance its manufacturing capabilities and achieve significant advancements in technological innovation. The goals of the "Made in China 2025" plan are multifaceted. These include increasing the domestic content of core components and materials to 40% by 2020 and 70% by 2025, reducing reliance on foreign technology, improving production efficiency, and promoting sustainable

manufacturing practices (Wübbeke et all., 2016, p.7). Additionally, the plan aims to elevate China's global standing in manufacturing quality and innovation, ensuring that Chinese products are recognized for their technological sophistication and reliability.

However, according to the Chinese lawyer Interviewee 2- the first document that marked the beginning of the real changes was the "New Generation Artificial Intelligence Development Plan", issued by China in July 2017. The plan focuses on fostering a robust AI ecosystem that includes research and development, industrial applications, talent cultivation, and ethical standards (Webster et al., 2017). Specifically, it targets advancements in key areas such as machine learning, cognitive computing, and intelligent robotics. The emphasis of the document is on the integration of AI with various industries to enhance productivity and drive economic growth. It encourages the development of AI technologies that can be applied in manufacturing, healthcare, agriculture, and urban management (Webster et al., 2017). The plan also highlights the importance of building an AI-friendly infrastructure, including high-speed data networks and advanced computing facilities. According to the Interviewee 2, the "New Generation Artificial Intelligence Development Plan" serves as the foundational stage for implementing AI regulations in China. It sets national standards for AI development and application, ensuring that AI technologies are developed responsibly and ethically. The plan also initiates research on the legal issues surrounding AI, addressing potential risks and establishing a legal framework to govern AI use. Interviewee 2 emphasizes that this proactive approach to regulation underscores China's commitment to creating a balanced environment where AI can thrive while safeguarding public interests and individual rights.

Interviewee 2 stated that the People's Republic of China (PRC) Cybersecurity Law, enacted by the Standing Committee of the National People's Congress in November 2016, and implemented in June 2017 aimed at enhancing cybersecurity measures and safeguarding data privacy within the country. One key aspect of the PRC Cybersecurity Law relevant to AI regulation is the requirement for network operators to store certain data within China's borders (Creemers et al., 2021). This provision aligns with the government's objective of maintaining control over data flows and ensuring data sovereignty. Additionally, the law empowers Chinese authorities to conduct spot-checks on companies' network operations, reinforcing regulatory oversight in the cybersecurity domain. Moreover, the law imposes mandatory testing and certification of computer equipment for critical sector network operators (Creemers et al., 2021). These requirements, outlined in Article 21 of the law,

emphasize the importance of internal security management systems, network security protections, and measures to prevent cyber attacks (Creemers et al., 2021). According to the Interviewee 2 - network operators are also mandated to implement data classification, backup procedures for important data, and encryption protocols, reflecting international best practices for data protection. In the context of AI regulation, the PRC Cybersecurity Law provides a foundational framework for ensuring the security and integrity of AI systems and the data they process. By requiring stringent cybersecurity measures and data localization practices, the law contributes to the establishment of a secure environment for AI development and deployment in China. Additionally, the law's emphasis on network security aligns with broader efforts to mitigate cybersecurity risks associated with AI technologies, further enhancing trust and confidence in AI systems among users and stakeholders (Creemers et al., 2021).

In June 2016, 46 companies from the U.S., Europe, and Japan signed a letter to Premier Li Keqiang, criticizing the cybersecurity law and claiming that it would hinder the entry of foreign companies and stifle innovation (Martina M., 2016). The law requires operators to store data within China and provide it to authorities for spot checks. This has raised concerns among some foreign companies about increased data control and heightened risks of intellectual property theft. The Chinese government has the right to request source code, encryption, and other strategic information, which could seriously harm companies' competitiveness if this data falls into the hands of competitors (Creemers et al., 2021).

Moreover, Article 9 of the law states that "Network operators carrying out business and service activities must follow laws and administrative regulations, respect social morality, abide by commercial ethics, be honest and credible, perform obligations to protect cybersecurity, accept supervision from the government and public, and bear social responsibility" (Creemers et al., 2021). The vagueness of these provisions allows the government to expand the grounds for inspections and reduces the ability of foreign companies to challenge demands for access to critical information (Wagner, 2017).

According to legal experts, the law supports domestic companies, as inspections can be conducted at the request of the government or trade associations, allowing Chinese companies to obtain strategically important data from their competitors. To ensure data localization, foreign companies will either have to invest in new data servers in China,

subjecting them to spot checks, or hire local server providers, thereby supporting the development of Chinese technology through service payments.

Jack Wagner, an analyst for Asia at PGI Intelligence, believes the law is aimed at bringing data under Chinese jurisdiction, making it easier to prosecute individuals who violate Chinese law (Wagner, 2017).

The 2018 White Paper on AI Standardization underscores the importance of addressing privacy concerns amid the rapid development of AI technologies. It highlights the need to protect personal data as AI systems increasingly analyze and utilize vast amounts of sensitive information (Triolo & Ding, 2018). Emphasizing the principle of human interests, the paper advocates for AI technologies to benefit human welfare while ensuring liability and consistency of rights and responsibilities (section 3.3.2).

Furthermore, the white paper acknowledges the convenience and potential benefits of government agencies collecting and using citizens' personal data for personalized services:

"*In addition, the development of AI technology makes the government's collection and use of citizens' personal data more convenient. Large amounts of personal data can help government departments to better understand the status of the people they serve and guarantee the opportunity and quality of personalized service. However, it follows that the risks and potential harms of improper use of personal data by government departments and government workers should be given sufficient attention*." (Section 3.3.3, Triolo & Ding, 2018)

Presented in Beijing in April 2018, the White Paper was introduced to the first meeting of SC 42, a standards committee with significant global influence. Led by a large Chinese delegation comprising government representatives and leading private sector companies, the meeting established four working groups focused on foundational standards, computational approaches, trustworthiness, and use cases and applications.

In addition to the White Paper, the China Association of Artificial Intelligence (CAAI) developed ethical guidelines aimed at guiding the responsible development and deployment of AI technologies. These guidelines were established to address growing concerns surrounding the ethical implications of AI and to promote the adoption of ethical practices within the AI industry. The CAAI's ethical guidelines cover various aspects of AI

development and usage, including data privacy, transparency, accountability, fairness, and safety. They emphasize the importance of respecting user privacy rights and ensuring the protection of personal data throughout the AI lifecycle. Transparency is also highlighted as a key principle, urging AI developers and providers to be open and clear about how their systems operate and make decisions. Moreover, the guidelines underscore the importance of accountability, urging AI developers to take responsibility for the outcomes of their technologies and to mitigate any potential harms they may cause. Fairness in AI algorithms and decision-making processes is emphasized to prevent bias and discrimination, ensuring that AI systems treat all individuals equitably.

With rapid development of technologies over time ethical guidelines were developed and strengthened throughout the legal acts issued by key technological actors. For example, in 2021, the Ministry of Science and Technology in China published a Code of Ethics for the New Generation of Artificial Intelligence (AI). The Code of Ethics for the New Generation of AI builds upon previous ethical guidelines and initiatives, incorporating updated principles and standards to address emerging challenges and developments in AI technology. One of the central tenets of the code is the emphasis on human-centric AI development, highlighting the importance of AI technologies benefiting society and individuals while respecting human rights and dignity. It underscores the need for AI systems to prioritize the well-being and interests of human beings, promoting inclusivity, diversity, and equality in AI applications. Transparency and accountability are also key components of the code, with provisions calling for AI systems to be transparent in their operations and decision-making processes. Developers and providers are urged to provide clear explanations of how AI algorithms work and how decisions are made, enabling users to understand and evaluate the outcomes of AI technologies. Furthermore, the code emphasizes the importance of fairness and non-discrimination in AI systems, urging developers to mitigate bias and ensure equitable treatment of all individuals. It calls for AI technologies to be designed and implemented in a manner that respects cultural diversity and avoids reinforcing existing inequalities. Safety and security considerations are also addressed in the code, with provisions aimed at ensuring the reliability, robustness, and security of AI systems. Developers are encouraged to implement measures to prevent AI-related accidents and mitigate risks associated with malicious exploitation of AI technologies.

In 2021, China implemented a comprehensive set of regulations aimed at governing various aspects of artificial intelligence (AI) technology. These regulations include the Data Security Law enacted in June 2021, the Personal Information Protection Law issued in November 2021, and the Critical Information Infrastructure Security Protection Regulations introduced in September 2021. The Data Security Law focuses on safeguarding data-related aspects of AI technology. It sets forth provisions to ensure the secure handling, storage, and transmission of data, including data used in AI systems. By establishing legal frameworks for data governance and security, this law aims to prevent data breaches, unauthorized access, and misuse of data in AI applications.

Following the Data Security Law, the Personal Information Protection Law establishes guidelines for the collection, processing, and utilization of personal data, including data used to train and operate AI systems. It requires organizations to obtain consent from individuals before collecting their personal information and mandates measures to protect the privacy and rights of data subjects in AI-related activities.

In addition to these laws focused on data and personal information, China also implemented the Critical Information Infrastructure Security Protection Regulations in September 2021. These regulations aim to ensure the security and resilience of critical information infrastructure (CII) against cyber threats, including those posed by AI technologies. By defining requirements for the protection of CII assets and systems, these regulations contribute to the overall cybersecurity framework governing AI applications in critical sectors.

The Cyber Administration of China (CAC) serves as the primary regulatory authority overseeing artificial intelligence (AI) and related technologies in China. As the central government agency responsible for internet governance and cybersecurity, the CAC plays a crucial role in formulating and implementing policies, regulations. The CAC monitors compliance with AI-related laws, regulations, and standards and enforces them through inspections, investigations, and enforcement actions. It works to ensure that AI companies and organizations adhere to the established rules and guidelines and take appropriate measures to address any violations or breaches. Additionally, the CAC collaborates with other government agencies, industry associations, research institutions, and international organizations to coordinate AI-related initiatives and promote cooperation in areas such as data sharing, technology development, and cybersecurity.

The Regulation on the Management of Algorithmic Recommendations in the Information Services of the Internet, issued in China in 2021, aims to govern the use of algorithmic recommendation systems by internet information service providers. These recommendation systems play a significant role in influencing user behavior and preferences by suggesting content, products, or services based on user data and behavior patterns. However, some scholars saw the misuse of this law by the government by making operators of algorithms follow ethical code for "development of positive energy" in the Internet and spread of unwilling or unlawful information.

Filipova (2024) highlights the significant implications of the Regulation on the Management of Algorithmic Recommendations as it serves to address various concerns in the realm of artificial intelligence (AI) and information services, aiming to protect user rights and ensure fair and transparent practices (Filipova, p.51). One crucial aspect of the regulation is its focus on curbing monopolistic practices by AI platforms, thereby safeguarding user rights (Filipova, p.51). It mandates special attention to the needs of elderly users to prevent fraud and prohibits the algorithmic generation of fake news, as well as the use of discriminatory or biased user tags in recommendation systems (Filipova, p.51). Furthermore, the regulation imposes obligations on service providers to prevent discriminatory and exploitative conditions for platform workers, such as couriers and drivers (Filipova, p.51). These measures aim to ensure fair treatment and working conditions for individuals involved in platform-based services. Additionally, service providers are required to work towards reducing the spread of harmful information and granting consumers the right to disable algorithmic recommendations (Filipova, p.51). This empowers users to control their online experiences and mitigate potential negative impacts of algorithmic recommendations on their interests.

However, according to the Interviewee 2 this law can be misused- Article 9 focuses on strengthening content management by creating a database to identify illegal and undesirable information, and improving the standards, rules, and procedures for database entries. It requires immediate action to halt the dissemination of illegal information once discovered and to prevent its further spread, with incidents being recorded and reported to the Cyber Administration of China (CAC).

The year 2022 saw the introduction of the Provisions on the Administration of Deep Synthesis Internet Information Services in China. This regulatory framework brought about

the emergence of municipal acts aimed at governing specific aspects of internet information services at a local level. These municipal acts, stemming from the broader provisions, are designed to address the complexities and nuances of deep synthesis internet information services within specific geographic jurisdictions. They outline guidelines, rules, and requirements for the administration and operation of such services, taking into account local needs, conditions, and priorities.

In 2023, China introduced the Regulation on the Interim Measures for the Management of Generative AI Services aimed to address the potential risks associated with generative AI services, which have the capability to create highly realistic synthetic media, including images, videos, and audio recordings. Key provisions of the regulation included measures to monitor and regulate the creation and dissemination of synthetic media to prevent misuse for malicious purposes such as disinformation, propaganda, or other activities harmful to national security or public order. It also mandated that providers of generative AI services implement safeguards to protect against unauthorized use and abuse of these technologies. Furthermore, the regulation established mechanisms for oversight and enforcement, empowering regulatory authorities to monitor compliance with the interim measures and take appropriate action against violations. This included measures such as inspections, audits, and penalties for non-compliance. The CAC's Generative AI Measures highlight that the Chinese government prioritizes information control and oversight (Halm et al., 2023). Article 4 of these measures explicitly states that generative AI services must align with socialist core values and avoid producing anti-government content (Halm et al., 2023). Additionally, providers of generative AI services with the potential to influence public opinion or mobilize society must comply with security assessments and filing requirements as outlined in China's "Internet Information Service Algorithmic Recommendation Management Provisions"(Halm et al., 2023).

In 2023, China introduced the Regulation on Facial Recognition Technologies, aiming to address the ethical and privacy concerns associated with the widespread use of facial recognition technology in various sectors. This regulation represents a significant step toward regulating the deployment and usage of facial recognition technologies to protect individuals' rights and interests. Key provisions of the regulation include strict guidelines on the collection, storage, and use of facial data, emphasizing the importance of obtaining informed consent from individuals before capturing or processing their facial images. It also imposes limitations on the purposes for which facial recognition technology can be employed,

prohibiting its use for discriminatory or unlawful activities. Moreover, the regulation mandates the implementation of safeguards to prevent the misuse or abuse of facial recognition technology, including measures to ensure data security and prevent unauthorized access to facial data. It also requires transparency and accountability from organizations deploying facial recognition systems, requiring them to disclose information about the purposes, methods, and implications of their facial recognition practices.Additionally, the regulation establishes mechanisms for oversight and enforcement, empowering regulatory authorities to monitor compliance with the regulatory requirements and take corrective action against violators. This may include penalties for non-compliance and sanctions against entities found to be engaging in unethical or unlawful uses of facial recognition technology.

The Regulation on Facial Recognition Technologies, enacted in 2023, serves the critical purpose of governing the utilization of facial recognition systems, with a dual focus on safeguarding personal information and maintaining public order and security (p.53). It introduces stringent measures to curtail the unauthorized and discriminatory use of facial recognition technology, explicitly prohibiting its application for analyzing sensitive attributes such as race, ethnicity, religious beliefs, and health status, except under specific circumstances such as obtaining consent or ensuring national security (p.53). While law enforcement and state security agencies have traditionally employed facial recognition extensively, the new regulations extend restrictions to non-state entities, including banks, airports, and hotels, aiming to mitigate potential privacy violations and abuses (p.53).

Furthermore, the regulations prescribe security protocols mandating the installation of image collection and identification equipment in public areas to enhance security measures. However, facility managers are not compelled to exclusively rely on facial recognition for entry and exit control; they must offer alternative identity verification methods to individuals who opt out of facial recognition (p.53). This provision underscores the regulation's commitment to balancing technological advancements with individual privacy rights and ensuring accessibility for all individuals, thereby fostering public trust and confidence in the responsible deployment of facial recognition technologies.

In 2023, the Chinese Academy of Social Sciences unveiled a significant law draft aimed at serving as a model legislation concerning the benefits of artificial intelligence (AI) for society. This comprehensive law draft emphasizes several key principles, including

prioritizing the welfare of the people, ensuring safety, openness, transparency, legal responsibility, fairness, equality, resource efficiency, and environmental protection. The overarching goal is to promote innovation, foster international cooperation, and uphold legality and legitimacy in the development and deployment of AI technologies.

By prioritizing the well-being of individuals and communities, the law draft underscores the importance of AI technologies in enhancing societal welfare and addressing critical challenges. Moreover, by emphasizing safety, openness, and transparency, the legislation aims to build trust and confidence among stakeholders while mitigating risks associated with AI deployment. Legal responsibility provisions seek to hold AI developers and operators accountable for their actions, ensuring that they adhere to ethical standards and comply with regulatory requirements.

The draft advocates for fairness and equality in AI applications, striving to prevent discrimination and promote inclusivity in access to AI-driven services and opportunities. Additionally, it underscores the importance of resource efficiency and environmental protection, aligning with broader sustainability goals. Moreover, the law draft underscores the imperative of promoting innovation in AI and fostering international cooperation to leverage shared expertise and resources. By emphasizing legality and legitimacy, the legislation seeks to establish clear guidelines and frameworks for AI development and deployment that respect legal norms and societal values.

Although China can be considered as a leader in AI regulation, there are legislative shortcomings. According to Filipova (2024) duplication of norms in various normative legal acts and fragmented regulation (p.58) can be viewed as a significant flaw of Chinese AI regulation. Additionally, according to the Filipova the extensive use of facial recognition technologies in China serves to maintain order and decrease common crime rates, but it also encroaches on privacy rights and enables the social ranking of citizens (p.58). This system uses AI technologies to automate the tracking of offenses, assigning points for each detected violation (p.58). Accumulating more points results in a lower social ranking, which can limit access to jobs, services, travel, and other benefits (p.58). Therefore, the regulation of AI in China prioritizes social control and the projection of state power over the protection of individual rights, justifying the suppression of personal freedoms for public interests.

According to Filipova (2024): “Thus, the presence of ideology does not exclude the corrective impact of public discourse on the formation of ethical regulation, under the influence of which legal regulation is built” (p.59). However, this opinion is controversial and in the next part of this chapter we will see that China is emphasizing social control over personal freedoms, political regime and ideology from the regulation and the governmental approach towards the emerging development of the artificial intellect.

3.3 Use of AI by Chinese government to promote national interests

The main organization which is responsible for the rulemaking, administrative licensing and punishment activities in Chinese cyberspace is Cyberspace Administration of China (CAC). Different experts consider this organization not only the main regulatory body but also a strong political decision-maker. According to the Dr. Rogier Creemers, assistant professor in modern Chinese studies at Leiden University- CAC has several different roles: First of all, CAC has regulatory control over all online content in China, including AI-generated content and recommendation algorithms (Kuo, M. A., 2023). This makes it the primary rule-setter for the world's largest online population and second-largest digital economy. Under the Personal Information Protection Law and the Data Security Law, the CAC is responsible for protecting personal information and handling data security issues (Kuo, M. A., 2023). The CAC directly oversees several important technical organizations, including: China’s DNS registry (CNNIC), China’s computer emergency response team (CNCERT/CC), the cybersecurity standardization body (TC260) and Cybersecurity Association of China (Kuo, M. A., 2023). Additionally, CAC houses the secretariat of the central cybersecurity and Informatization Commission: top-level decision-making body, chaired by Xi Jinping, includes leaders from key party and state bodies, the People’s Bank of China, and the military (Kuo, M. A., 2023). The CAC runs this secretariat, providing crucial information to the Commission and implementing its policy decisions (Kuo, M. A., 2023). This direct link to top leadership underscores its significant political influence.

A Law professor from the Hong Kong University, Angela Huyue Zhang in her book argues that The Cyberspace Administration of China (CAC), closely linked with the propaganda bureau of the Communist Party, is responsible for minimizing the risk of politically harmful content being generated by AI models (Zhang, A. H., 2024). Some of the restrictive measures include the mandatory alignment of language models with socialist values and the requirement for user identification (Zhang, A. H., 2024). Chinese chatbots, such as "Spark"

(星火) from iFlytek, are developed based on large language models (LLM) similar to Western counterparts like ChatGPT. However, they are trained on unique datasets that include built-in safety measures to prevent the generation of politically sensitive content. One of the key censorship goals in China is to prevent so-called "historical nihilism", which refers to information and commentary that contradict the official ideological line of the Chinese Communist Party (CCP) (Green et al., 2024). Authorities use algorithmic keyword detection and manual moderation to control information about historical events, such as the Great Leap Forward, the Cultural Revolution, and the events at Tiananmen Square (Green et al., 2024). Special hotlines are created to combat historical nihilism, and social media companies are required to prevent the spread of such content (Green et al., 2024).

In the case of China, there is limited information on how AI is used for domestic propaganda, but there is abundant information on international cyberattacks. The Microsoft Threat Intelligence 2024 report highlights various activities and threats originating from Chinese-based entities, emphasizing the extensive use of AI technologies in cyber-espionage, influence operations, and misinformation campaigns on the international arena.

Among key threats and activities highlighted in the report are espionage, influence operations using AI, operations targeting critical infrastructure, AI- generated news and deep fakes, conspiratorial narratives and misinformation.

According to the report, during the summer of 2023, the China-based espionage group Gingham Typhoon targeted nearly every South Pacific Island country, focusing on international organizations, government entities, and the IT sector. This group employed sophisticated phishing campaigns, affecting even critics of the Chinese government and diplomatic allies of China (Microsoft, 2024, p.4). Additionally, Chinese threat groups, particularly Storm-0062, significantly increased their activities in late 2023, targeting U.S. military entities and critical infrastructure. This included compromising defense-related government entities and contractors involved in aerospace, defense, and natural resources critical to U.S. national security (Microsoft, 2024,p.6).

The most successful cyber group in China was the Storm-1376 also known as Spamouflage and Dragonbridge. According to the report "*Ai and Cover Influence Operations: Latest Trends*" Storm-1376 is characterized as: "Persistent Chinese threat actor posting content across the internet to praise China and criticize its critics" (OpenAI, 2024, p.23). According

to the report: “Some of this operation’s social media activity was attributed by the FBI to a unit within China’s Ministry of Public Security” (OpenAI, 2024, p.24). Storm-1376 actively uses AI in order to conduct sophisticated influence operations, spreading misinformation, and shaping public opinion across various regions and topics. Throughout 2023, AI-generated memes (funny pictures on the internet) targeted the United States, amplifying controversial domestic issues and criticizing the current administration. These operations have been consistent in leveraging AI to enhance the volume and frequency of influence campaigns (Microsoft, 2024, p.6). It was the first case where Microsoft Threat Intelligence identified how an acting figure of the national government used AI-generated content to influence results of international elections. Storm-1376 has been using AI-generated news anchors and enhanced videos in campaigns targeting Taiwan and Myanmar. These campaigns include false depictions and inflammatory remarks aimed at discrediting political figures and influencing public opinion (Microsoft, 2024, p.7). Storm-1376 has been active in spreading conspiratorial content around high-profile geopolitical events. Notable campaigns included misleading narratives about U.S. government actions, such as falsely claiming that wildfires in Hawaii were caused by a "weather weapon" (Microsoft, 2024, p.9). This group also criticized Japan's disposal of nuclear wastewater, spreading doubt about the safety of the process and accusing the U.S. of malicious intentions (Microsoft, 2024, p.10). Storm-1376's activities extended to South Korea, where they amplified protests against Japan's wastewater disposal and criticized the South Korean government. The group also spread misinformation about a train derailment in Kentucky, promoting anti-U.S. government conspiracy theories and political division (Microsoft, 2024, p.10).

Another significant aspect of Chinese use of the AI technologies is highlighted in the report from OpenAI published in May 2024. Utilizing AI models, the network performed tasks such as debugging code, conducting social media analysis, and generating research content. One notable example from 2023 includes the use of AI models to set up the WordPress theme for revealscum[.]com, a website that published derogatory content about Chinese government critics, labeling them as "traitors" (OpenAI, 2024, p.24). The network also employed AI for open-source research, such as obtaining information on applying for developer accounts on social media and summarizing posts by Chinese government critics. However, AI-generated content constituted only a fraction of the overall output. Many identified accounts on platforms like X and Medium produced a higher volume of manually created content in both

English and Chinese. The mixture of AI-generated and manually created content helped maintain a degree of authenticity and avoided detection (OpenAI, 2024, p.26).

China employs various tools to exert social control, including social surveillance combined with social scoring systems.

China's deployment of facial recognition technology is extensive, encompassing a network of cameras that surveil nearly every aspect of urban life. These cameras capture and digitize individuals' movements, not only for access control purposes but also for identifying and even publicly shaming individuals engaged in minor infractions such as jaywalking (Kaur, 2023). This surveillance infrastructure extends to databases like that of SenseNets Technology, which, when leaked in 2019, revealed the extent to which personal data, including sensitive information about Uighur individuals, is collected and utilized by Chinese authorities (Kaur, 2023).

Since the enactment of new anti-terrorism laws in 2014, China has increasingly relied on AI and machine learning algorithms to monitor and regulate its population (Miracola, 2019). The Integrated Joint Operations Platform (IJOP) is pivotal in this extensive surveillance framework, notably deployed in Xinjiang to oversee the predominantly Uyghur and Turkic Muslim populations (Wang, 2019). The IJOP integrates data from multiple sources such as gas stations, street checkpoints, and controlled areas like communities and schools. Using CCTV camera feeds, it aggregates these data points into a centralized platform for continuous monitoring and analysis (Wang, 2019). This technology aims to minimize human involvement in security operations, although police officials remain crucial for data collection (Miracola, 2019). IJOP, used in Xinjiang to send Uyghurs to the Chinese established "reeducation and training" camps and it is a prominent example of AI-driven social control (Miracola, 2019). These camps detain Uyghurs and impose indoctrination with Communist Party propaganda, discouraging Islamic practices, and encouraging adherence to Han Chinese cultural norms (Miracola, 2019). The IJOP app enhances surveillance capabilities beyond mere data collection by alerting officials to suspicious activities. It prompts immediate investigations, including searches of individuals' phones for prohibited software, network tools, or content like VPNs used to bypass government internet restrictions (Wang, 2019).

The Chinese government's embrace of facial recognition has been a cornerstone of its surveillance strategy, enabling efficient monitoring and control of public spaces and

individual behavior. However, recent developments indicate a regulatory shift aimed at curbing potential abuses of this technology. The Cyberspace Administration of China (CAC) released draft regulations in August 2023 to govern facial recognition technology, emphasizing protections for personal information and limitations on discriminatory uses such as racial or ethnic profiling (Kaur, 2023).

Beyond facial recognition, China has implemented social scoring systems that leverage AI and big data analytics to assess and rank individuals based on their behavior and adherence to societal norms. These systems compile data from various sources, including financial transactions, social media activity on platforms like WeChat, and interactions with government services (Strittmatter, 2021). Through algorithms, individuals are assigned scores that influence access to social privileges such as loans, travel permissions, and even job opportunities (Strittmatter, 2021).

AI plays a critical role in these scoring systems by automating the analysis of vast datasets to evaluate citizens' social creditworthiness. Its ultimate objective goes beyond traditional notions of effective governance to include the regulation of citizen conduct and the enhancement of their ethical standards, responsibilities typically attributed to the government (Benedetto, A. , 2023, p.9). This automated process enables real-time monitoring and assessment of behavior, facilitating proactive intervention by authorities when deviations from expected norms occur (Strittmatter, 2021).

In conclusion, China's strategic integration of AI across regulatory, geopolitical, and domestic surveillance domains underscores its dual use as both a tool for national security and social control. While advancing technological capabilities contribute to economic growth and cybersecurity resilience, they also pose significant ethical and geopolitical challenges. As AI continues to evolve, balancing innovation with accountability remains crucial to mitigating risks and safeguarding democratic values on a global scale.

Conclusion:

Artificial intelligence (AI) has emerged as a pivotal technology, influencing various aspects of modern life and prompting nations to develop regulatory frameworks to manage its integration and impact. Scholars such as Cath, Floridi, and Taddeo and others have highlighted the significance of AI development and the emphasis governments worldwide place on these technologies.

The EU AI Act is a comprehensive, horizontal regulation that categorizes AI systems into four risk levels: unacceptable risk, high risk, limited risk, and minimal risk. This framework aims to protect fundamental rights and promote ethical AI development through stringent requirements for high-risk applications, including transparency, human oversight, and robustness.

The advantages of this approach include the prioritization of ethical standards and human rights, ensuring AI development aligns with democratic values. The Act promotes trust in AI systems through stringent transparency and accountability measures, and its risk-based approach allows for targeted regulation, minimizing barriers for low-risk AI applications while ensuring high-risk applications are rigorously controlled.

However, the comprehensive nature of the regulation can lead to bureaucratic hurdles, potentially slowing innovation. High compliance costs for high-risk AI systems and different interpretations across member states could lead to inconsistencies and market fragmentation.

Russia's AI regulatory approach is less codified and more experimental, relying on Experimental Legal Regimes (ELRs) to test and refine AI technologies. The focus is on state security and control, with systems like Roskomnadzor's "Vepr" and "Oculus" used for surveillance and public opinion monitoring. The advantages of this approach include rapid innovation, as the flexible, experimental approach allows for swift development and implementation of AI technologies. Centralized control enables uniform standards and quick policy adjustments, and the emphasis on state security supports national stability and

protection. However, extensive surveillance raises significant privacy and civil liberties issues, and the focus on state interests over individual rights can lead to ethical concerns. The lack of a unified framework may result in legal ambiguities and inconsistencies, and the fragmented and experimental nature of AI regulation may lead to inconsistent protections and unpredictable legal environments.

China employs a sector-specific, vertical approach to AI regulation, with numerous targeted regulations addressing different aspects of AI, such as recommendation algorithms and facial recognition. The focus is on leveraging AI for national and public goods while maintaining state control. The advantages of this approach include strong state support and a focus on innovation, which facilitate quick advancements in AI technology. AI is a key component of China's national development goals, promoting economic growth and social benefits. The sector-specific approach allows for tailored regulations that address specific issues effectively. However, extensive use of AI for surveillance and social control raises significant ethical and human rights issues. The prioritization of state interests can undermine individual freedoms and public trust, and wide discretion for public authorities can lead to arbitrary enforcement and lack of accountability.

The regulatory differences between the EU, Russia, and China have profound societal implications.

In the EU, the emphasis on ethical AI and human rights aims to build public trust and ensure AI technologies contribute positively to society. The stringent requirements for high-risk AI systems are designed to protect individuals from potential abuses and ensure that AI development aligns with democratic values. In Russia, the focus on state security and control through AI technologies contributes to a surveillance state, potentially undermining individual freedoms and fostering mistrust among citizens. The fragmented and experimental nature of AI regulation may lead to inconsistent protections and unpredictable legal environments. In China, the regulatory framework supports rapid innovation but prioritizes state control, potentially stifling broader societal benefits and individual freedoms. The focus on public goods and national benefits can enhance societal acceptance of AI, despite privacy and ethical concerns.

The differing regulatory frameworks for artificial intelligence (AI) in the European Union (EU), Russia, and China reflect and influence profound geopolitical dynamics.

The EU's comprehensive and ethically focused AI regulation positions it as a global leader in promoting responsible AI. This regulatory model enhances the EU’s diplomatic leverage and soft power in international forums. By championing robust ethical standards and human rights protections in AI development, the EU not only sets a precedent but also influences global norms, shaping the trajectory of AI governance worldwide. Despite not being the leader in AI technology innovation, the EU leverages its regulatory prowess to assert its values on the global stage, evident in regulations like the GDPR and the recently enacted AI Act. These efforts are critical for asserting digital sovereignty and safeguarding the EU market against external influences.

Conversely, Russia’s approach to AI regulation underscores its broader geopolitical strategy of maintaining domestic control and exerting influence on the international stage. Russia employs AI primarily for state surveillance and propaganda purposes, aligning with its internal security imperatives and external political maneuvers. The emphasis on surveillance technologies such as "Oculus" and "Vepr" reflects Russia's strategic prioritization of maintaining domestic stability and suppressing dissent, often at the expense of individual privacy and civil liberties. This approach not only exacerbates tensions with countries advocating for human rights and ethical AI practices but also shapes Russia's image and relationships in the global arena.

In contrast, China’s rapid development and export of AI technologies are central to its geopolitical ambitions. China sees AI as a critical tool for achieving global leadership in technology and innovation, prioritizing economic growth and national security over individual rights and freedoms. Despite international criticisms concerning human rights abuses and surveillance practices facilitated by AI, China continues to expand its influence through strategic investments in digital infrastructure abroad. Internationally, China leverages AI for cyber-espionage and influence operations, utilizing entities like Storm-1376 to advance its geopolitical interests through hybrid warfare tactics. Domestically, AI technologies such as the Integrated Joint Operations Platform (IJOP) in Xinjiang exemplify China's deployment of AI for social surveillance and repression, underpinning authoritarian governance practices and human rights violations.

Looking ahead, while China refines its AI ethics guidelines and promotes international cooperation on responsible AI development, debates persist about the balance between state control and individual freedoms within its regulatory framework. These geopolitical dynamics underscore the global implications of AI regulation, highlighting how differing approaches shape international relations, influence technological standards, and define norms of governance in the digital age.

The comparative analysis of AI regulations in the EU, Russia, and China reveals distinct approaches shaped by their governance structures and strategic goals. The EU's comprehensive and ethical framework aims to balance innovation with the protection of human rights, reflecting its commitment to democratic values and digital sovereignty. Russia's experimental and security-focused approach prioritizes state control, reflecting its geopolitical strategy and internal challenges. China's sector-specific, innovation-driven framework promotes rapid development while maintaining state oversight, aligning with its national development goals. These regulatory frameworks not only shape the societal impacts of AI within these regions but also influence global power dynamics and international relations. As AI continues to evolve, understanding these differences and their implications is crucial for navigating the complex landscape of AI governance and its role in global strategy.

Bibliography

A.Alexandrov, S. Korsakov (2021, February 9) Вычислить по камерам и привлечь к ответственности. Как в России выслеживают участников протестов [Identify through cameras and hold accountable." How participants in protests are tracked down in Russia], Current time, Retrieved from:https://www.currenttime.tv/a/russia-detentions-navalny/31092461.html

Agentstvo (2023, February 8) Роскомнадзор занялся созданием бот-фермы [Roskomnadzor set up a botfarm] Retrieved from: https://www.agents.media/rkn-bot-ferma/

AI Association to Draft Ethics Guidelines. Chinese Academy of Sciences. (2019, January 10). https://english.cas.cn/print_2019/index.shtml?docurl=https%3A%2F%2Fenglish.cas.cn%2Fnewsroom%2Fnews%2F201901%2Ft20190110_203885.shtml

AlgorithmWatch, ‘EU Parliament votes on AI Act; member states will have to plug surveillance loopholes’: https://algorithmwatch.org/en/eu-parliament-votes-on-ai-act/

Alliance for Artificial Intelligence (2023) Code of Ethics for Artificial Intelligence. Retrieved from: https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf

Almada, M., & Petit, N. (2023). The EU AI Act: Between product safety and fundamental rights. SSRN. Retrieved from : https://www.researchgate.net/profile/Marco-Almada/publication/366464842_The_EU_AI_Act_Between_Product_Safety_and_Fundamental_Rights/links/63a2c27641663a23c073ffb2/The-EU-AI-Act-Between-Product-Safety-and-Fundamental-Rights.pdf

Amnesty International: ‘EU: Artificial Intelligence rulebook fails to stop proliferation of abusive technologies’: https://www.amnesty.org/en/latest/news/2024/03/eu-artificial-intelligence-rulebook-fails-to-stop-proliferation-of-abusive-technologies/

Bashlykova, N., & Krylova, E. (2024, May 28) Голосовые связи: за дипфейки хотят ввести уголовную ответственность [Dip fakes want to criminalize], Izvesia, Retrieved from: https://iz.ru/1702846/natalia-bashlykova-elizaveta-krylova/golosovye-sviazi-za-dipfeiki-khotiat-vvesti-ugolovnuiu-otvetstvennost

Baturov, T. (2023, October 23). Суды стали чаще назначать реальные сроки за «фейки» и дискредитацию армии [Courts have begun to impose actual sentences more frequently for "fake news" and discrediting the army]. Forbes. Retrieved from: https://www.forbes.ru/society/498992-sudy-stali-case-naznacat-real-nye-sroki-za-fejki-i-diskreditaciu-armii

BBC News. (2023, February 9). Архитектура российской цензуры: что мы узнали из крупнейшей утечки в истории Роскомнадзора. Главное [Architecture of Russian censorship: what we have learned from the biggest leak in the history of Roskomnadzor. The most important].BBC News. Retrieved from https://www.bbc.com/russian/features-64576925

Benedetto, A. (2023). The assessment impact of Artificial Intelligence (AI) on basic human rights: the Chinese case study.

Biber, S. E. (2021). Machines Learning the Rule of Law. Verfassungsblog: On Matters Constitutional.

Bill № 512628-8 On amendments to the Federal Law "On experimental legal regimes in the sphere of digital innovations in the Russian Federation" (on improving the implementation of pilot legal regimes for digital innovation) Retrieved from: https://sozd.duma.gov.ru/bill/512628-8

Brattberg, E., Rugova, V., & Csernatoni, R. (2020). Europe and AI: Leading, lagging behind, or carving its own way? (Vol. 9). Washington, DC, USA: Carnegie endowment for international peace.

Castets-Renard, C., & Besse, P. (2022). Ex Ante accountability of the AI Act: Between certification and standardization, in pursuit of fundamental rights in the country of compliance. Pursuit of Fundamental Rights in the Country of Compliance (August 29, 2022). Artificial Intelligence Law: Between Sectoral Rules and Comprehensive Regime. Comparative Law Perspectives, C. Castets-Renard & J. Eynard (eds), Bruylant, Forthcoming.

Central Bank of Russian Federation (2023) Application of Artificial Intelligence on the financial market. Report for public consultation. Retrieved from: https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf

Council of Europe. (1950). European Convention on Human Rights. Retrieved 2024, from https://eur-lex.europa.eu/EN/legal-content/glossary/european-convention-on-human-rights-echr.html

Creemers , R., Webster , G., & Triolo, P. (2021, September 1). Translation: Cybersecurity law of the People's Republic of China (effective June 1, 2017). DigiChina. https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china-effective-june-1-2017/

Criminal Code of the Russian Federation Article 207.3. Public Dissemination of Knowingly False Information about the Use of the Armed Forces of the Russian Federation, the Exercise of Powers by State Bodies of the Russian Federation, or the Provision of Assistance by Volunteer Formations, Organizations, or Individuals in the Performance of Tasks Assigned to the Armed Forces of the Russian Federation or the National Guard Troops of the Russian Federation. Retrieved from: https://www.consultant.ru/document/cons_doc_LAW_10699/19bf2b8e4b62e143a17a50041a204252d0e263ce/

Criminal Code of the Russian Federation Article 280.3. Public Actions Aimed at Discrediting the Use of the Armed Forces of the Russian Federation for the Purpose of Protecting the Interests of the Russian Federation and its Citizens, Maintaining International Peace and Security, the Exercise of Powers by State Bodies of the Russian Federation, or the Provision of Assistance by Volunteer Formations, Organizations, or Individuals in the Performance of Tasks Assigned to the Armed Forces of the Russian Federation or the National Guard Troops of the Russian Federation Retrieved from: https://www.consultant.ru/document/cons_doc_LAW_10699/1aa9268e7d3bd57bcbd46a3016641c5af64b9c87/

Critical information infrastructure security protection regulations (effective Sept. 1, 2021). DigiChina. (2021, October 14). https://digichina.stanford.edu/work/translation-critical-information-infrastructure-security-protection-regulations-effective-sept-1-2021/

Decree of the Government of the Russian Federation dated 19 August 2020 2129-r On approval of the Concept of Development of Regulation of Relations in the Sphere of Artificial Intelligence and Robotics for the period up to 2024. Retrieved from: https://www.garant.ru/products/ipo/prime/doc/74460628/

Decree of the President of the Russian Federation from 10.10.2019 №490 "On the development of artificial intelligence in the Russian Federation" (together with the "National Strategy for the development of artificial intelligence for the period up to 2030"). Retrieved from: https://a-ai.ru/wp-content/uploads/2024/03/%D0%9D%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B5%D0%B3%D0%B8%D1%8F_%D1%80%D0%B0%D0%B7%D0%B2%D0%B8%D1%82%D0%B8%D1%8F_%D0%98%D0%98_2024.pdf

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. Retrieved from: https://uobrep.openrepository.com/bitstream/handle/10547/623613/1_s2.0_S026840121930917X_main.pdf?sequence=4

Ebers, M., Hoch, V. R., Rosenkranz, F., Ruschemeier, H., & Steinrötter, B. (2021). The European Commission's proposal for an Artificial Intelligence Act—a critical assessment by members of the Robotics and AI Law Society (RAILS). J, 4(4), 589-603.

Ethical norms for New Generation Artificial Intelligence released. Center for Security and Emerging Technology. (2021, October 21). https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

European Commission. (2018). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions: Artificial Intelligence for Europe (COM/2018/237 final). Retrieved 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN

European Commission. (2018, April 25). Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines [Press release]. Brussels. Retrieved from: https://ec.europa.eu/commission/presscorner/detail/en/IP_18_3362

European Commission. (2020, February 19). WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust (COM(2020) 65 final). Brussels. Retrieved 2024, from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065

European Disability Forum: ‘AI Act agreement: partial win on accessibility’: https://www.edf-feph.org/ai-act-agreement-partial-win-on-accessibility/

European Parliament, & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council 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). Retrieved 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

European Parliament, & Council of the European Union. (2019). Directive (EU) 2019/790 of the European Parliament and of the Council on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. Retrieved 2024, from https://eur-lex.europa.eu/eli/dir/2019/790/oj

European Parliament. (2024, February 2). Provisional agreement resulting from interinstitutional negotiations: Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.

Committee on the Internal Market and Consumer Protection, Committee on Civil Liberties, Justice and Home Affairs. Retrieved from : https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf

European Union. (2012). Charter of Fundamental Rights of the European Union. Official Journal of the European Union, C 326, 391–407. Retrieved [insert date], from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

Federal Law "On Amendments to the Federal Law "On Information, Information Technologies and Information Protection" and Separate Legislative Acts of the Russian Federation on the Regulation of the Exchange of Information through Information and Telecommunication Networks" Retrieved from: https://ww.consultant.ru/document//cons_doc_LAW_162586/

Federal Law "On Conducting Experiment on Establishment of Special Regulation to Create Necessary Conditions for Development and Introduction of Artificial Intelligence Technologies in the Subject of the Russian Federation - city of federal importance Moscow and amendments to articles 6 and 10 of the Federal Law "On Personal Data" of 24.04.2020 №123-FZ Retrieved from: https://www.consultant.ru/document/cons_doc_LAW_351127/

Federal Law "On Experimental Legal Regimes in the Sphere of Digital Innovations in the Russian Federation" from 31.07.2020 N 258-FZ Retrieved from: https://www.consultant.ru/document/cons_doc_LAW_358738/

Federal Law "On Personal Data" of 27.07.2006 №152-FZ Retrieved from:https://base.garant.ru/12148567/

Federal Law of 02.07.2021 N 331-FZ «On the introduction of amendments to individual legislative acts of the Russian Federation in connection with the adoption of the Federal Law «On experimental legal regimes in the sphere of digital innovations in the Russian Federation». Retrieved from: https://www.consultant.ru/document/cons_doc_LAW_389015/

Federal State Unitary Enterprise «Main Radio Frequency Center» (2019). УСТАВ ПРОЕКТА «Кабинет оперативного взаимодействия» [CHARTER OF THE PROJECT Cabinet of operational interaction»]. Federal State Unitary Enterprise «Main Radio Frequency Center» Retrieved from: https://static.istories.media/uploaded/documents/2b35fee588d5438a819ddd1b20ac642c.pdf

Federal State Unitary Enterprise «Main Radio Frequency Center» (2020). Утвержденная программа «От единого информационного пространства к единой ведомственной цифровой платформе» [Approved program «From a single information space to a single departmental digital platform»] Main Radio Frequency Center Retrieved from:https://rkn.gov.ru/docs/UTVERZHDENNAJA_VPCT_RKN_2021-2023_sajt_11.pdf

Filipova, I. A. (2024). Legal Regulation of Artificial Intelligence: Experience of China. Journal of Digital Technologies and Law, 2(1), 46-73.

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. Philosophy & Technology, 36(1), 15.

Future of Life Institute (2022). General purpose AI and the Ai Act, https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf

Gavriluk, A., & Rozhkov, R. (2023, February 27). Смотрите свою речь: насколько эффективен мониторинг Роскомнадзором российского интернета. [Watch your speech: How effective can Roskomnadzor's monitoring of the Russian Internet be]. Forbes.Retrieved from: https://www.forbes.ru/tekhnologii/485333-sledi-za-rec-u-naskol-ko-mozet-byt-effektiven-monitoring-runeta-ot-roskomnadzora

Green, K., Miller, F., Faerber , G., Henry, M., Wise, H., Sprott, A. S., Lafferty , Dr. B., & Francis, E. (2024, February 20). Censorship practices of the People’s Republic of China. Censorship Practices of the People’s Republic of China .

https://www.uscc.gov/sites/default/files/2024-02/Censorship_Practices_of_the_Peoples_Republic_of_China.pdf

Gstrein, O. J. (2022). European AI Regulation: Brussels Effect versus Human Dignity?. Zeitschrift für Europarechtliche Studien (ZEuS), 4. Retrieved from: https://research.rug.nl/files/261303007/ZEuS_4_2022_2._Umbruch_Gstrein.pdf

Halm, K. C., Wu, W., Seiver, J., & Austin, P. J. (2023, July 28). China's Cyberspace Administration releases "interim" rules regulating the use of Generative AI. Davis Wright Tremaine.
https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/07/china-issues-generative-ai-regulations

Joint statement by Access Now, AlgoRights, Amnesty International, Article 19, EDRi, Irish Council for Civil Liberties, European Disability Forum, AlgorithmWatch, Bits of Freedom, Fair Trials, LEFÖD, European Center for Not-for-Profit Law, Kermes, and Politiscope. (2021, December 2). EU's AI Act fails to set gold standard for human rights. Retrieved from https://algorithmwatch.org/en/ai-act-fails-to-set-gold-standard-for-human-rights/

Joint statement by ECNL, Liberties, & ECF; Day, J., Iwańska, K., Simon, E., & Willamo, K. (2024, April). Packed with loopholes: Why the AI Act fails to protect civic space and the rule of law. Civil Liberties Union for Europe e.V. https://www.liberties.eu
Kaur, D. (2023, August 14). Is facial recognition use to be curbed by Law in China?. TechHQ.
https://techhq.com/2023/08/why-is-china-is-changing-its-stance-on-facial-recognition-after-decades-of-surveillance/

Kornya, A. (2023, October 23). Полная ответственность: Ужесточены уголовные наказания за дискредитацию армии, а административные наказания сократились. [Full extent of responsibility: Criminal penalties for discrediting the army have increased, while administrative ones have decreased]. Kommersant, (197/P), 3. Retrieved from: https://www.kommersant.ru/doc/6295801

Kuo, M. A. (2023, November 28). In a league of its own: The Cyberspace Administration of China. – The Diplomat.
https://thediplomat.com/2023/11/in-a-league-of-its-own-the-cyberspace-administration-of-china/

Kuteynikov, D., Izhaev, O., Lebedev, V., & Zenin, S. (2022). Legal regulation of artificial intelligence and robotic systems: review of key approaches. Cuestiones Políticas, 40(72), 690-703.

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk.

Regulation & Governance, 18(1), 3-32. Retrieved from: https://onlinelibrary.wiley.com/doi/pdf/10.1111/rego.12512

Lyubimov, A. (2024, June 15). Political persecution under Putin highest since Stalin era – proekt. The Moscow Times. https://www.themoscowtimes.com/2024/02/23/political-persecution-under-putin-highest-since-stalin-era-proekt-a84218

Martina, M. (2016, August 11). Business groups petition china's premier on cyber rules.

Reuters. Retrieved from https://www.reuters.com/article/idUSKCN10M1DN/.
Microsoft . (2024). (issue brief). Same targets, new playbooks: East Asia threat actors employ unique methods (pp. 1–16). Retrieved 2024, from https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/MTAC-East-Asia-Report.pdf.

Miracola, S. (2019, June 4). How China uses A.I. to Control Society - ISPI. How China Uses Artificial Intelligence to Control Society. https://www.ispionline.it/sites/default/files/pubblicazioni/isp_commentary_miracola_04.06.2019.pdf

OpenAI. (2024, May). AI and covert influence operations: Latest trends. https://downloads.ctfassets.net/kftzwdyauwt9/5IMxzTmUclSOAcWUXbkVrK/3cfab518e6b10789ab8843bcca18b633/Threat_Intel_Report.pdf

Personal information protection law of the People's Republic of China. PIPL. (2022, May 10). https://personalinformationprotectionlaw.com/#:~:text=The%20PIPL%20came%20into%20effect,legal%20basis%20and%20disclosure%20requirements.
Polyakova, A., & Meserole, C. (2019). Exporting digital authoritarianism: The Russian and Chinese models. Policy Brief, Democracy and Disorder Series, 1-22.

Radio Liberty. (2022, November 19). Хакеры взломали внутреннюю сеть структуры Роскомнадзора [Hackers have hacked the internal network of the structure of Roskomnadzor]. Radio Liberty. Retrieved from https://ww.svoboda.org/a/hakery-vzlomali-vnutrennyu-setjroskomnadzora/32137862.html.

RosComFreedom (2018) Новгородская область призывает «Демона Лапласа» для слежки за пользователями [Novgorod region calls «Daemon Laplace» to spy on users] Retrieved from:https://roskomsvoboda.org/ru/post/novgorodskaya-oblast-prizyivaet-dem/

RosComFreedom (2021) Власти Самарской области проследят за детьми с «депрессивным и суицидальным поведением [Samara region authorities to monitor

children with depressive and suicidal behavior] Retrieved from: https://roskomsvoboda.org/ru/post/slezhka-za-suic-podrost-v-socsetyah/

Ruggiu, D. (2018) Human Rights and Emerging Technologies: Analysis and Perspectives in Europe. with the Prefation of Roger Brownsword, Pan Stanford Publishing, Singapore

Shaelou, S. L., & Razmetaeva, Y. (2024, January). Challenges to Fundamental Human Rights in the age of Artificial Intelligence Systems: shaping the digital legal order while upholding Rule of Law principles and European values. In ERA Forum (pp. 1-21). Berlin/Heidelberg: Springer Berlin Heidelberg. Retrieved from: https://link.springer.com/content/pdf/10.1007/s12027-023-00777-2.pdf

Smuha, N. A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., & Yeung, K. (2021). How the EU can achieve legally trustworthy AI: a response to the European Commission's proposal for an Artificial Intelligence Act. Available at SSRN 3899991. Retrieved from: https://strathprints.strath.ac.uk/85567/1/Smuha_etal_SSRN_2021_How_the_EU_can_achieve_legally_trustworthy_AI.pdf

Sokolova, M. (2024, April 19). Почему растет количество иностранных агентов, пояснили в Госдуме. [Why the number of foreign agents is growing, explained in the State Duma.] Parlamentskaya Gazeta. Retrieved from:https://www.pnp.ru/politics/v-gosdume-obyasnili-pochemu-rastet-kolichestvo-inoagentov.html

Strittmatter, K. (2020). We have been harmonized: Life in China's Surveillance State. HarperCollins.

SVTV(2023, 30 September) Роскомнадзор закупает бот-фермы для поиска запрещённой информации [Roskomnadzor purchases bot-farms for search of forbidden information]Retrieved from: https://svtv.org/news/2023-09-30/roskomnadzor-zakupaiet/

The National New Generation Artificial Intelligence Governance Specialist Committee (国家新一代人工智能治理专业委员会). (2021, October 21). Ethical norms for New Generation Artificial Intelligence released. Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

The National People's Congress of the People's Republic of China , Data Security Law №84 of the People's Republic of China (2021). Retrieved from http://www.npc.gov.cn/englishnpc/c2759/c23934/202112/t20211209_385109.html.

Triolo, P., & Ding, J. (2018, June 20). Translation: Excerpts from China's "White Paper on Artificial Intelligence Standardization." DigiChina. https://digichina.stanford.edu/work/translation-excerpts-from-chinas-white-paper-on-artificial-intelligence-standardization/

Tunyayeva M. (2023, February 13). Роскомнадзор запустил систему автоматического поиска запрещенного контента «Окулус».[Roskomnadzor launched a system of automatic search of prohibited content «Oculus»] Vedomosti. Retrieved from https://www.vedomosti.ru/technology/articles/2023/02/13/roskomnadzor-zapustil-sistemu-avtomaticheskogo-poiska-zapreschennogo-kontenta-okulus

Turkova, K. (2023, November 22). Семь лет за пять стикеров, приговор Саши Сколиченко [Seven years for five stickers: Sasha Skocilenko's sentence.] VOA News. Retrieved from: https://www.golosameriki.com/a/skochilenko_text/7364688.html

Ustinova A., Kinyanina E (2024, 11 January) МВД привлечет нейросети к поиску правонарушителей [Ministry of Internal Affairs will involve AI to search of offenders] Retrieved from: https://www.vedomosti.ru/technology/articles/2024/01/11/1014513-mvd-privlechet-neiroseti-k-poisku-pravonarushitelei

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97-112. Retrieved from: https://arxiv.org/pdf/2107.03721

Verstka (2023, December 15). Определяют ситуативно: Как людей и организации включают в списки иноагентов. [Determined situationally: How individuals and organizations are included in the lists of foreign agents] Retrieved from: https://verstka.media/kak-lyudey-i-organizacii-vkluchayut-v-spiski-inogentov
Wagner , J. (2017, June 1). China's cybersecurity law: what you need to know. The Diplomat . Retrieved from https://thediplomat.com/2017/06/chinas-cybersecurity-law-what-you-need-to-know/.

Wang, M. (2019, October 11). Interview: China's "big brother" app. Human Rights Watch. https://www.hrw.org/news/2019/05/01/interview-chinas-big-brother-app

Webster, G., Creemers, R., Kania , E., & Triolo, P. (2017, August 1). Full translation: China's "new generation artificial intelligence development plan" (2017). DigiChina. https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

Wörsdörfer, M. (2024). Mitigating the adverse effects of AI with the European Union's artificial intelligence act: Hype or hope?. Global Business and Organizational Excellence,

43(3), 106-126. Retrieved from: https://www.researchgate.net/profile/Manuel-Woersdoerfer/publication/375577655_Mitigating_the_Adverse_Effects_of_AI_With_the_European_Union's_Artificial_Intelligence_Act_Hype_or_Hope/links/655002fbb86a1d521bd662d0/Mitigating-the-Adverse-Effects-of-AI-With-the-European-Unions-Artificial-Intelligence-Act-Hype-or-Hope.pdf

Wu, F., Lu, C., Zhu, M., Chen, H., Zhu, J., Yu, K., ... & Pan, Y. (2020). Towards a new generation of artificial intelligence in China. Nature Machine Intelligence, 2(6), 312-316.

Wübbeke, J., Meissner, M., Zenglein, M. J., Ives, J., & Conrad, B. (2016). Made in china 2025. Mercator Institute for China Studies. Papers on China, 2(74), 4.

Zhang, A. H. (2024). High Wire: How China Regulates Big Tech and Governs Its Economy. Oxford University Press.

互联网信息服务算法推荐管理规定. 互联网信息服务算法推荐管理规定_中央网络安全和信息化委员会办公室. (2021, November 16). https://www.cac.gov.cn/2022-01/04/c_16428946063