Ali Karami [1]

Afshin Mottaghi Dastenaei [2]

Artificial Intelligence Governance in the European Union

Journal of Electrical Systems

Abstract

The European Union has taken an active stance in regulating AI through legislative initiatives, ethical guidelines, and consultation with stakeholders (European institutions, companies and governments) and aims to lead the global AI landscape by balancing AI advances with the social values of citizens. Through AI governance, the Union seeks to create a regulatory framework that promotes innovation while ensuring the protection of fundamental rights, safety, and the public interest. This article aims to examine the EU mechanisms for AI governance. The main research question is what are the most important salient features of the AI governance model in the EU? Given the qualitative nature of the subject, the research method is descriptive-analytical. The required data will be collected by referring to EU AI documents and laws such as the AI Law, the General Data Protection Regulation, ethical guidelines, the European AI Strategy, the Data Governance Law, the AI Coordination Plan, and the EU AI Supervisory and Enforcement Mechanisms. The paper's findings show that while the EU is progressing in AI governance, it faces challenges in harmonizing member states' regulations, attracting global cooperation, navigating the rapidly evolving AI landscape, and balancing innovation with ethical standards. The paper's findings show that the EU AI governance model focuses on creating a regulatory framework that encourages innovation and oversees the development of AI so that it is used in a way that is consistent with European values such as protecting fundamental rights, ensuring transparency, and promoting accountability.

Keywords: Artificial Intelligence, Governance, European Union, Supervision, Innovation, Ethical Guidelines.

INTRODUCTION

As AI continues to evolve, the need for consistent and forward-looking governance has increased, making it a cornerstone of the responsible advancement of this technology. AI governance refers to the frameworks, policies, and principles that guide the ethical development, deployment, and use of AI technologies. As AI is increasingly integrated into various sectors—from healthcare and finance to defense and entertainment—the need for strong governance structures increases because AI systems can make decisions that have important consequences for individuals and societies (Mottaghi Dastenaei & Karami, 2024). Without clear governance, AI can perpetuate biases, violate privacy, or make unethical decisions. Proper governance ensures that the technology operates within ethical guidelines that prioritize fairness, justice, and respect for human rights. AI can inherit biases from the data it is trained on, leading to discriminatory outcomes in areas such as hiring, law enforcement, or credit scoring. Governance establishes mechanisms to detect and correct bias. Clear rules about who is responsible for decisions made by AI systems (developers, users, or the AI itself) are critical to preventing harm and ensuring accountability (Almeida et al. 2021).

AI is associated with technical and societal risks, ranging from security vulnerabilities (such as AI systems being hacked) to unintended consequences in decision-making (such as autonomous vehicles causing accidents). Governance helps mitigate these risks through regulation, oversight, and risk management strategies. Governance frameworks encourage transparency in AI development. Implementing safety protocols and ensuring continuous monitoring of AI systems helps prevent catastrophic failures, especially in critical industries such as healthcare or transportation (McLean et al. 2021: 649–663).

The debate on AI governance stems from a variety of disciplines and reflects different perspectives on how AI should be regulated and managed. From a technologically optimistic perspective, AI is a powerful tool that, if properly managed, can solve major societal problems and drive innovation across industries. Proponents emphasize that AI governance should prioritize innovation while maintaining ethical concerns. Proponents of this view believe that AI governance should not stifle innovation with overly restrictive policies. This view suggests that industries developing AI systems are better placed to regulate themselves, with ethical guidelines coming from within organizations rather than imposed from outside. AI is seen as potentially driving economic growth,

[1] .**Corresponding author :**Alikarami598@ut.ac.ir
[2] .**Email addres:** A.mottaghi@khu.ac.ir
PhD in European Studies, University of Tehran
Full Professor of Political Geography, Kharazmi University

increasing productivity, and creating new jobs. From this perspective, governance should focus on maximizing these economic benefits while addressing risks such as job displacement (Danaher, 2022: 1-29).

The human rights perspective, with its focus on ethics, emphasizes the importance of grounding AI governance in human rights principles and moral values, arguing that AI poses significant risks to privacy, autonomy, and social equality if left unchecked. AI systems can exacerbate existing biases in society. Proponents of this perspective emphasize the importance of governance frameworks that prevent AI from exacerbating discrimination in areas such as law enforcement, employment, or lending. Governments should prioritize protecting individuals' data from abuse. This perspective supports regulations such as the European Union's General Data Protection Regulation, which restricts how companies can collect, store, and use personal data. Ethical AI frameworks often seek "explainable AI," where algorithms are transparent in their decision-making process, and for accountability structures that ensure that humans remain accountable for AI actions (Hunkenschroer & Kriebitz, 2023: 199–213).

From a social justice perspective, AI governance is not just about mitigating risks or promoting innovation, but also about ensuring that AI development and its benefits are fairly distributed. Without careful governance, AI can reinforce existing inequalities. There are concerns that AI may widen the wealth gap with large corporations and wealthy countries benefiting disproportionately, while low-income communities and developing countries may be left behind. Governments should address these inequalities through inclusive policies. Advocates of this view argue for policies that democratize access to AI technology so that marginalized groups and developing countries can also reap its potential benefits. As AI automates tasks, there are concerns about job displacement, particularly in sectors that employ low-wage workers. Governments should ensure that workers are supported through retraining programs and social safety nets (Zaidan & Ibrahim, 2024).

The national security and defense perspective focuses on the risks that AI poses to national security and global stability. Proponents are concerned about the role of AI in warfare, cybersecurity, and surveillance. There are ongoing debates about the ethical implications of using AI in military applications, particularly autonomous weapons systems that can make lethal decisions without human intervention. Governance in this context is about setting international laws and conventions to prevent misuse. AI systems can become targets for cyberattacks or, worse, weapons for cyberwarfare. Governance must therefore include strong standards to protect AI systems from external threats. Governments can use AI for mass surveillance, which raises concerns about privacy and civil liberties. From this perspective, governance must include strict regulations on how AI can be used for surveillance and law enforcement (Johnson, 2019: 147-169).

From global governance and international cooperation perspective, the global reach of AI means that a fragmented approach to governance can create conflict and unequal protections across countries. This perspective requires international cooperation to create standard AI governance frameworks. Advocates argue that governance frameworks should be global in nature and establish unified standards across borders that ensure fairness, safety, and interoperability of AI systems. In the defense context, there is a risk that countries will compete to develop AI for military use, leading to an AI arms race. Global governance could help create agreements similar to the nuclear non-proliferation treaties to prevent this. Cooperation between countries, research institutions, and technology companies is essential to create a globally accepted ethical standard for AI development (Tallberg et al. 2023: 1–23).

From an environmental sustainability perspective, AI governance also intersects with environmental sustainability concerns. From this perspective, AI governance should consider the environmental impact of AI systems, particularly energy consumption and their potential to contribute to climate change. AI systems, especially those involving large-scale machine learning models, require significant computational resources. Proponents of this perspective argue for governance frameworks that promote energy-efficient AI systems and encourage sustainable computing practices. AI can be a powerful tool to combat climate change by optimizing energy systems, reducing waste, and increasing environmental stewardship. Governments should promote the use of AI for sustainability initiatives while minimizing its carbon footprint (the amount of greenhouse gases emitted) (van Wynsberghe, 2021: 213-218).

Proponents of the futurist view argue that AI governance should prepare for the possibility of superintelligent AI that could surpass human intelligence and autonomy. They believe that the risks of such AI systems could be existential and require both immediate and long-term management strategies. From this perspective, AI

governance should prioritize preventing the development of AI systems that could evolve beyond human control and potentially threaten humanity. The focus is on ensuring that AI systems are aligned with human values and goals. This view suggests that failure to align could lead to AI pursuing goals that conflict with human well-being. Governance frameworks should adopt the precautionary principle, which suggests that AI development should proceed with caution, especially when it comes to advanced forms of AI. Different perspectives on AI governance reflect different priorities, whether fostering innovation, protecting human rights, ensuring equitable access, or protecting against existential risks. These perspectives are not mutually exclusive, and effective AI governance may require a multifaceted approach that integrates ethical, economic, security, and international considerations to ensure the responsible development of AI for the benefit of all (Coeckelbergh, 2024).

THEORETICAL APPROACH

Several major theories of AI governance focus on how to regulate, manage, and monitor the development and use of this technology. These theories often draw from existing legal, ethical, and governance frameworks while proposing new ideas tailored to the unique challenges posed by AI;

1. Soft Law and Self-Regulation Theory

This theory advocates a flexible, non-binding approach to AI governance, allowing industries to create and enforce their ethical guidelines without heavy government intervention. Its proponents believe that technological innovations evolve at a rate that traditional government regulations cannot keep up with and that industries are best equipped to create standards specific to their AI applications. Soft law includes codes of conduct, industry standards, and voluntary frameworks rather than binding legal regulations. For example, technology companies may adopt ethical principles such as fairness, accountability, and transparency on their own. Self-regulation could be more adaptable than government regulation, which may be too slow to respond to evolving AI technologies. However opponents of the theory argue that self-regulation may lack accountability and enforcement, potentially allowing unethical AI practices to flourish, especially when commercial interests are at stake (Marchant & Gutierrez, 2023: 376–424).

2. co-regulation initiatives

Participatory theory proposes a hybrid model in which industry self-regulation is combined with government oversight. This approach ensures flexibility for technological innovation while providing external scrutiny to ensure compliance with ethical standards and the prevention of harm. In this model, industries develop guidelines in consultation with government agencies, and governments take action to enforce compliance when necessary. Governments can collaborate with private companies to create standard frameworks that ensure the ethical use of AI, sharing responsibilities between regulators and innovators. For example, the European Union's approach to regulating emerging technologies, such as the AI Act, balances industry input with government enforcement and represents a co-regulatory model (de Almeida et al. 2021: 5050–525) .

3. Precautionary Principle Theory

The precautionary principle is a widely recognized approach to environmental governance that can be applied to AI. It proposes that AI development should proceed with caution, especially when the risks and long-term consequences of some AI technologies are uncertain or potentially harmful. The core of the theory is that it is better to prevent potential harm than to react to it after it has occurred. In the case of AI, this means that high-risk AI systems should be subject to stricter governance measures, even if the full range of risks is not known (Munthe, 2020).

4. Agile Governance Theory

Agile governance emphasizes adaptation and continuous learning in AI policy. The theory advocates for governance frameworks that evolve as quickly as the technology itself, allowing governments and regulatory bodies to respond to the dynamic nature of AI. Rather than fixed regulations that may become outdated, agile governance encourages iterative and flexible policies that are regularly reviewed. This allows for continuous updating as AI technology advances. Agile governance calls for collaboration between governments, industry managers, technologists, ethicists, and civil society organizations to co-shape policies. The theory supports the

idea of regulatory sandboxes that allow AI innovators to test new technologies in a controlled environment under government supervision before widespread deployment (Almeida et al. 2021).

5. Ethical Governance Theory

This theory argues that AI governance should be fundamentally based on ethical principles to ensure fairness, transparency, and accountability. It emphasizes the creation of regulatory frameworks based on core human values such as dignity, justice, and equality. AI systems should be aligned with ethical values and human rights, focusing on fairness, explainability, accountability, and avoiding bias or discrimination. Organizations such as the European Union and companies such as Google have developed AI ethical guidelines that promote the responsible use of AI. These focus on principles such as transparency (the ability to explain AI decisions) and non-abuse (ensuring that AI does not cause harm). While widely supported, ethical governance alone may not be enforceable without legal structures. Ethical codes can be vaguely enforced and open to interpretation (Camilleri, 2024:1-15).

6. Risk-based governance theory

This theory suggests that AI governance should be tailored to the level of risk that different AI applications pose. Instead of uniform regulation, AI systems are regulated based on their potential harm or societal impact. Different AI systems pose different risks and the governance model should reflect this. For example, low-risk AI applications (such as product recommendation algorithms) will face minimal oversight, while high-risk AI (such as autonomous driving or AI in healthcare) will be heavily regulated. Under this approach, AI systems are categorized into risk levels, and governance frameworks are designed based on those categories. The proposed EU AI law follows this model by identifying "high-risk" AI systems that require strict regulation. This approach can underestimate risks in areas where the impact of AI is not immediately apparent, such as long-term privacy concerns or subtle social biases introduced by AI into everyday applications (Novelli, et al. 2024).

7. Public Governance Theory

This theory places the public interest and citizen participation at the center of AI governance and emphasizes democratic decision-making, transparency, and the inclusion of community voices in shaping AI policies. Governance frameworks should be created with active public input and ensure that AI systems reflect societal values and norms. The theory focuses on participatory governance, where civil society organizations, NGOs, and ordinary citizens participate in AI policymaking. AI decisions that affect people should be transparent, and there should be mechanisms for public feedback and influence on AI deployment, particularly in government services or public infrastructure such as smart cities. Large-scale public participation in complex technical areas such as AI may be difficult and can slow down policymaking processes. Furthermore, non-experts may struggle to meaningfully participate in highly technical discussions (Züger & Asghari, 2023: 815-828.).

8. The theory of technocratic governance

This theory advocates that AI governance should be left to technical experts and specialists who have a deep knowledge of AI systems and their potential implications. AI is a highly specialized and technical field and should be governed by those who understand the complexities of the technology. This model suggests that governments should defer to panels of experts, technologists, and industry leaders to shape AI regulations. Regulatory bodies are largely comprised of engineers, data scientists, AI ethicists, and other professionals who create policies based on their knowledge of the risks and capabilities of AI. Critics argue that this can lead to technocratic elitism, where decisions about AI are made without sufficient public input, leading to a lack of transparency and responsiveness to broader societal concerns.

9. Global Governance Theory

This theory proposes that given the international reach and impact of AI technologies, AI governance should be conducted at the global level. Proponents of this theory argue that fragmented, national-level governance frameworks can lead to uneven support and competitive imbalances, necessitating global coordination. The borderless nature of AI requires international cooperation to establish global standards that ensure the ethical, safe, and equitable use of AI technologies worldwide. Proponents advocate the establishment of global institutions—similar to the United Nations or the World Health Organization—that would oversee AI regulation, establish global standards, and foster cross-border cooperation. For example, the Global Partnership on AI and other

international initiatives seek to promote collaboration between governments, industry, and civil society to develop common approaches to AI governance. Achieving global governance may be difficult due to differing political, cultural, and economic priorities across countries. Some worry that global standards could reflect the interests of powerful countries and companies while neglecting smaller countries or marginalized groups (Veale et al. 2023).

10. Embedded AI Governance Theory

This theory advocates governance by design, where regulatory frameworks and ethical safeguards are built directly into AI systems rather than externally imposed. AI systems themselves should be designed with governance mechanisms built into them to ensure that ethical principles are implemented at the technology level. This can include programming transparency, fairness, and accountability directly in algorithms. A central tenet of this theory is the development of explainable AI, in which the performance of AI systems is transparent, understandable, and interpretable by humans. Critics argue that while governance by design can prevent certain harms, it may not address the broader societal implications of AI deployment, such as labor market disruptions or regulatory concerns. AI governance theories reflect a diversity of approaches to managing the development and deployment of AI technologies. From precautionary principles that focus on preventing harm to public-centered approaches that emphasize inclusiveness, these theories offer different frameworks for ensuring that AI serves society ethically and safely. As AI continues to evolve, it may be necessary to integrate multiple theories of governance to address the complex and multifaceted nature of this transformative technology (Pflanzer et al. 2023: 1267–1271).

AI Governance Models Around the World

Different regions and countries have developed different models of AI governance, reflecting diverse approaches to regulation, innovation, ethics, and societal impacts. Each model reflects local legal frameworks, ethical priorities, and economic goals, with some regions focusing more on innovation and others prioritizing risk management. Here are the key AI governance models around the world:

1. The US Model: Innovation and Market-Driven

The US adopts a market-based and early AI governance model that focuses on promoting innovation, economic growth, and maintaining leadership in AI technology. The US approach generally involves sector-specific regulations rather than comprehensive and overarching laws. The US lacks a single national AI law (Alhosani & Alhashmi, 2024). Instead, AI is regulated through laws and agencies based on specific sectors, such as healthcare, finance, and transportation (for example, the Food and Drug Administration regulates AI in medical devices). In 2020, the White House Office of Science and Technology Policy issued AI Regulatory Principles that focus on ensuring AI innovation while managing risks. These principles encourage non-prescriptive and flexible regulatory approaches with a focus on self-regulation by the private sector. US technology companies (e.g., Google, Microsoft, and IBM) have developed internal AI ethics guidelines focused on fairness, transparency, and responsible AI, but these guidelines are voluntary and non-binding (Alhosani & Alhashmi, 2024). The US government supports significant research in AI through initiatives such as the National AI R&D Strategic Plan, which aims to maintain US leadership in AI technology. The US model, however, promotes innovation and economic competition but has been criticized for lacking strong regulatory safeguards against the risks of AI, such as bias, discrimination, and privacy violations. This approach places considerable reliance on the private sector to self-regulate and innovate (Donlon, 2024: 6-14).

2. The Chinese Model: State-led, Innovation-driven with Top-down Control

China's AI governance model is characterized by state-led innovation and strong government involvement in shaping AI development. With this model, China aims to become a global leader in AI by 2030, combining technological ambitions with top-down control. In 2017, the Chinese government issued the Next Generation AI Development Plan, setting the country's goal to lead the world in AI by 2030. The government provides significant funding and support for AI research and development. The country widely uses AI in surveillance, facial recognition, and social credit systems to monitor and control its population (Gong et al. 2024). AI technologies have been deployed to enhance state security and enforce social order. The country has issued some AI-specific regulations, such as laws on autonomous vehicles, facial recognition, and deepfakes. However, many other aspects

of governance are intertwined with other existing regulations on cybersecurity, privacy, and data protection. In 2021, China published AI ethics guidelines that focus on principles such as fairness, human-centered design, and ensuring that AI does not pose social risks. However, these ethical concerns are often secondary to the goals of state control and innovation. China's AI governance model prioritizes rapid innovation and economic dominance while using AI for state purposes. The focus on state control and oversight raises global concerns about privacy, human rights, and the export of China's AI governance model to other countries (Roberts et al. 2021: 59–77).

3. The UK Model: Innovation and Regulatory Flexibility

The UK has adopted an innovation-friendly AI governance model that aims to balance technological advances with appropriate safeguards for society and focuses on regulatory flexibility and leveraging existing laws. The UK National AI Strategy (2021) aims to position the UK as a global leader in AI by promoting research, innovation, and responsible AI. The strategy encourages businesses and regulators to work together on the ethics and governance of AI (Science, Innovation and Technology Committee, 2024).

The UK government supports low-friction regulation, aiming to minimize barriers to innovators while ensuring the safety and ethicality of AI systems. AI is regulated through existing legislation such as the Data Protection Act and Competition Act, without specific new legislation. The AI Council and the Centre for Data Ethics and Innovation play a key role in guiding AI governance by advising the government on AI policy, ethics, and societal impacts. The UK uses regulatory sandboxes – controlled environments where companies can test AI technologies with oversight to encourage innovation while reducing risks. Overall, the UK model promotes innovation and market development while ensuring that AI systems meet ethical and safety standards. The model emphasizes collaboration between the private sector and government but may face challenges in creating robust AI-specific regulations as the technology advances (Huw et al. 2023: 1–23).

4. The Canadian Model: Ethics and a Human Rights Approach

Canada's AI governance model emphasizes human rights, ethics, and responsible innovation. The country is known for its focus on the social and ethical implications of AI and has made efforts to incorporate human rights considerations into its AI governance framework. Canada was an early pioneer in AI ethics, establishing the Montreal Declaration for Responsible AI in 2018, which outlines principles for developing ethical AI, including fairness, accountability, and transparency. The Canadian government requires AI systems used in the public sector to undergo algorithmic impact assessments that assess the potential impact of AI systems on citizens and ensure accountability. The Pan-Canadian AI Strategy (2017) promotes AI research, innovation, and ethical considerations in AI development. Canada has established AI research centers, such as the Montreal Institute for Learning Algorithms, to develop AI (Attard-Frost et al. 2024).

AI governance is also managed at the provincial level, with jurisdictions such as Quebec and Ontario taking proactive steps to ensure that AI systems are aligned with human rights and social justice concerns. Canada's Ethics and Human Rights First approach demonstrates a commitment to responsible innovation. The emphasis on assessing algorithmic impact in the public sector represents a forward-looking governance model, but its implementation may be slower than market-based models.

5. The Japanese Model: Human-Centered AI Society

Japan's AI governance model is based on the vision of a super-intelligent society, where AI and other emerging technologies are integrated into society to create a human-centered and sustainable future. Japan emphasizes the role of AI in addressing societal challenges such as population aging and labor shortages. The country envisions a super-intelligent society in which AI technologies are seamlessly integrated to improve citizens' quality of life and address societal challenges. Japan's AI Technology Strategy (2017) emphasizes the development of AI that enhances human well-being, focusing on sectors such as healthcare, mobility, and public services. The country issued the AI Research and Development Guidelines for International Cooperation (2017) and the AI Utilization Guidelines (2019), which promote the development of AI in accordance with ethical principles such as human dignity, fairness, and privacy. The country supports global cooperation on AI governance and has worked with international organizations to develop common standards and ethical frameworks for AI. The Japanese model reflects a human-centered and socially responsible approach to AI governance that aims to integrate AI into all

aspects of life while addressing ethical concerns. Its focus on society 5 emphasizes the potential of AI to improve social well-being, but faces challenges in fostering AI innovation at the pace of countries such as the United States and China (Liu et al. 2024:105–121).

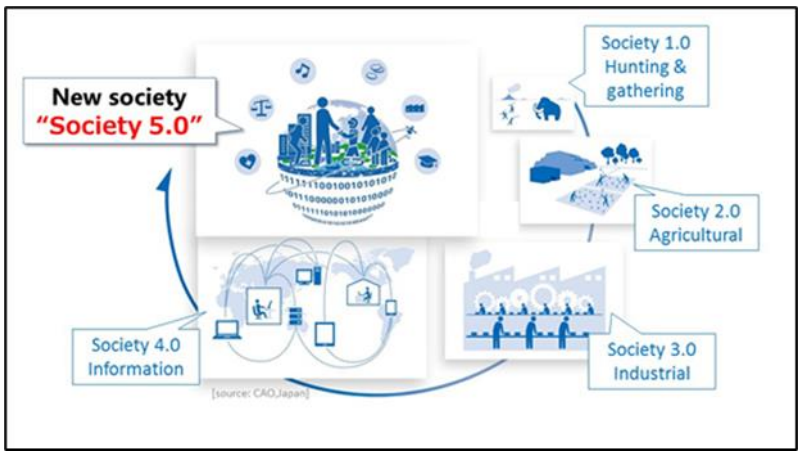

Figure 1. Japan's 5-Society Governance Model

Source: (https://www.wissenschaft-x.com)

These different models of AI governance—from the market-first innovation model in the United States to the authoritarian approach of the Chinese government—reflect a diversity of global perspectives. Each model has unique strengths and challenges, with the choice of model often dependent on the balance between fostering innovation and managing the risks associated with AI. As AI technology evolves, these governance frameworks are likely to adapt to increasing calls for international cooperation on global AI standards.

EU AI Documents and Legislation

The European Union has developed several important documents and legislation to guide AI governance. These documents outline the EU's approach to ensuring that AI development is aligned with ethical principles, fundamental rights, transparency, and risk management (Jobin et al. 389–399.).

The following is a review of the most prominent documents and legislation related to AI governance in the EU:

1. Artificial Intelligence Law 2021

The Artificial Intelligence Law is the cornerstone of the EU's AI governance framework and the first comprehensive legal framework for AI globally. The law aims to regulate the development, deployment and use of AI systems in the EU based on risk levels, ensuring that AI technologies are safe and respect human rights. The law classifies AI systems based on the risk to safety and fundamental rights, including:

1. Unacceptable risk: AI systems are completely banned (e.g. social scoring by governments).
2. High risk: Strict regulations for systems in sensitive areas (e.g. healthcare, law enforcement).
3. Limited risk: Transparency requirements for medium-risk systems (e.g. chatbots).
4. Minimal risk: Low-risk systems (e.g. video games) without specific regulations.

Highlights of the law include:

1. Transparency: High-risk AI systems must disclose key information about how they make decisions.
2. Human oversight: AI systems must allow for human intervention.
3. Data governance: High-risk systems require unbiased, high-quality datasets.
4. Fines: Companies face fines of up to 6% of their global revenue for non-compliance (EUR-Lex, 2024: https://eur-lex.europa.eu).

2. Coordinated Plan on Artificial Intelligence - 2021 Update

First adopted in 2018 and updated in 2021 to align with the objectives of the EU Digital Decade and the AI Act. The document aims to foster the development and deployment of AI across the EU, ensuring cooperation between Member States in research, innovation, and governance. The main focus areas of this document are:

1. Public sector AI: promoting the use of AI in healthcare, environment, and public administration.
2. AI research and innovation: supporting AI research with funding, in particular through the Horizon Europe program.
3. Re-skilling and upskilling: addressing the need for new skills in the AI workforce and promoting AI training.
4. Support: ensuring access to AI technologies and data resources for SMEs (Coordinated Plan on Artificial Intelligence, 2024: https://digital-strategy.ec.europa.eu/).

3. Ethical Guidelines for Trustworthy Artificial Intelligence - 2019

This document, prepared by a high-level expert group on artificial intelligence, sets out ethical principles for the development and use of artificial intelligence in the European Union. The document aims to guide the development of artificial intelligence in a way that supports human rights, democratic values , and social well-being.

The key ethical principles included in the document are:

1. Human agency and oversight: AI systems should empower individuals, not undermine human autonomy.
2. Technical robustness and safety: AI should be safe and trustworthy.
3. Privacy and data governance: AI should respect individuals' privacy and ensure proper data management.
4. Transparency: AI systems should be explainable and their performance should be understandable.
5. Diversity, non-discrimination and fairness: AI systems should not be biased or discriminatory.
6. Accountability: There should be clear mechanisms to hold AI systems accountable for their actions (Ethics guidelines for trustworthy AI, 2024: https://digital-strategy.ec.europa.eu/).

4. General Data Protection Regulation - 2018

Although not specifically an AI regulation, this document has a major impact on how personal data is processed by AI systems. It sets out strict requirements for privacy and data protection, which are particularly important for AI systems that manage large amounts of data.

Some of the provisions in this document that also apply to AI include:

1. Data protection by design: AI systems must be designed with data protection as a core principle.
2. Consent and transparency: Individuals must give informed consent for the use of their data by AI systems.
3. Right to explanation: In high-stakes decisions, individuals can challenge AI-based decisions and request an explanation of how the decision was made (General Data Protection Regulation, 2024: https://gdpr-info.eu/).

5. White Paper on Artificial Intelligence - 2020

The White Paper on Artificial Intelligence outlines the EU's vision for promoting trustworthy AI while ensuring safety and fundamental rights. It is one of the key documents that paved the way for the AI law. The document aims to promote the development of AI across the EU while managing potential risks for individuals and society. Some of the key proposals in the document are:

1. Regulatory framework: Proposes a risk-based framework for AI regulation, focusing on high-risk sectors.
2. Ecosystem of excellence: Proposes measures to ensure that Europe is at the forefront of AI innovation, including investment in research and skills development.
3. Ecosystem of trust: Establishes the need for an EU-wide governance structure to ensure the responsible development and use of AI (WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust, 2020: https://commission.europa.eu/).

6. European Strategy on Artificial Intelligence - 2018

The European Strategy on Artificial Intelligence outlines the EU's broader vision for the development of artificial intelligence, focusing on increasing Europe's global competitiveness and ensuring that artificial intelligence respects fundamental rights. The objectives of the document are:

1. Boosting investment in artificial intelligence: encouraging public and private investment in artificial intelligence research, in particular through Horizon Europe.
2. Addressing socio-economic change: preparing European societies and labor markets for the transformations that artificial intelligence will bring.
3. Ethical and human-centered artificial intelligence: ensuring that artificial intelligence is developed and used in a way that respects human dignity and democratic values (Jorge Ricart et al. 2022).

7. Cybersecurity Act - 2019

The Cybersecurity Act establishes a framework for EU-wide cybersecurity certification for products and services that are relevant to ensuring that artificial intelligence systems are safe and resilient. Under this law, a cybersecurity certification process can be applied to AI systems, especially in high-risk sectors such as healthcare, transport and financial services. AI systems deployed in critical infrastructure or sensitive areas must also meet strict cybersecurity standards. Taken together, these documents form a comprehensive approach to AI governance in the European Union, emphasizing ethics, safety, transparency, and data protection, while supporting innovation (Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity(

8. Digital Services Act

Enacted in 2022, this law regulates online platforms, including social media and e-commerce, moderates content and responsibilities of online platforms in managing illegal content and ensuring user safety. The Digital Services Act requires transparency about algorithms, including those used in AI-based recommendation systems. Some of the key principles of the law are:

1. Algorithmic transparency: Platforms must disclose how their algorithms work and the impact of their systems on users.
2. User empowerment: Users must be able to understand and control algorithmic decisions that affect them.

9. Digital Markets Act

Enacted in 2022, the law aims to ensure fair competition in digital markets and curb the power of big tech companies. The Digital Markets Act imposes obligations on large platforms (gatekeepers) regarding their use of AI in maintaining fair practices. Some of the main provisions of this law are:

1. Fairness and non-discrimination: Gatekeepers cannot unfairly abuse data or restrict access to services.
2. Interoperability: Encourages data sharing between platforms and the use of non-discriminatory algorithms (Regulation (EU) 2022/2065 of the European Parliament and of the Council).

EU AI Governance Model

The EU AI Governance Model is a proactive, risk-based and ethics-driven approach. It aims to ensure that AI technologies are safe, transparent and in line with fundamental human rights. The EU AI Governance Model seeks to strike a balance between innovation and regulation, focusing on fostering trust in AI systems while managing potential risks to society, individuals and democratic values (Justo-Hanani, 2022).

The components of the EU AI Governance Model are:

1. Risk-based approach

The core element of the EU AI Governance Model is the risk-based classification of AI systems, announced by the AI Law, a major legislative proposal. This approach categorizes AI systems based on their potential risk to safety, security and fundamental rights:

1. Unacceptable risk: AI systems that pose a serious threat to safety or human rights, such as social scoring by governments (such as China's social credit system), are strictly prohibited. These systems are considered too dangerous or harmful to be permitted in the EU.

2. High risk: AI systems operating in critical areas such as healthcare, law enforcement, education or employment are classified as high risk. These systems are permitted but face strict regulations. High-risk AI providers must meet strict requirements for transparency, human oversight, data quality and safety. Examples include AI in medical devices, recruitment tools and biometric identification systems.

3. Limited risk: AI systems with a medium potential for harm, such as chatbots or algorithms that personalise online content, must comply with transparency obligations. Users must be informed when interacting with AI systems to ensure they understand how their data is being used.

4. Minimal risk: AI systems with minimal or no risk, such as AI in video games or spam filters, face little regulatory intervention (Bildt, et al. 2020).

2. Artificial Intelligence Law

The Artificial Intelligence Law, proposed in April 2021, is the first comprehensive legal framework to regulate AI worldwide. It takes a risk-based approach and sets out rules for the development, use and deployment of AI across the EU. Some of the key aspects of the law include:

1. Transparency and accountability: High-risk AI systems must be transparent about how they make decisions and enable human oversight and accountability.

2. Data governance: High-risk AI systems must use high-quality datasets that are free from bias and ensure privacy.

3. Human oversight: The AI Law requires that high-risk AI systems remain under human oversight, ensuring that humans can intervene if necessary to prevent harm.

The AI Law also includes significant fines violations for; Fines of up to 6% of a company's annual global revenue demonstrate the EU's commitment to enforcing these rules (Brattberg et al. 2020).

3. Ethics and Trustworthy AI

The EU's AI governance model is strongly influenced by a strong commitment to ethical principles. In 2019, the European Commission published Ethical Guidelines for Trustworthy AI, which set out fundamental principles for AI systems that are aligned with human dignity, rights and well-being. The guidelines set out seven key requirements:

1. Human agency and oversight: AI should empower human decision-making and include oversight mechanisms.

2. Technical robustness and safety: AI systems should be resilient, secure and designed to minimise risks and harms.

3. Privacy and data governance: AI should respect user privacy and ensure proper data management.

4. Transparency: AI systems should be explainable and users should understand how the systems make decisions.

5. Diversity, non-discrimination and fairness: AI should be designed to prevent bias and ensure fairness for all users.

6. Social and environmental well-being: AI systems should benefit society and the environment and avoid harm.

7. Accountability: AI systems should have mechanisms to ensure accountability for actions and decisions.

8. These guidelines have strongly influenced the development of AI law and continue to shape AI governance across the European Union (Data Ethics Commission. 2019).

4. General Data Protection Regulation

The General Data Protection Regulation, which came into effect in 2018, plays a key role in the EU's AI governance model. Although not specifically an AI law, it governs how personal data is processed by AI systems

and sets strict standards for data privacy and security. The key requirements of the General Data Protection Regulation for AI are:

1. Data protection by design: AI systems must be built with privacy considerations at the core of their design.
2. Right to explanation: Although individuals are at issue, individuals can challenge AI-based decisions and seek explanations about how their data is processed, particularly in high-stakes decisions.
3. Consent and transparency: AI systems that process personal data must obtain user consent, ensure transparency in data processing, and protect personal information.

The General Data Protection Regulation helps protect individuals' privacy and ensures that AI systems using personal data comply with strong privacy standards (European Commission. (2018).

5. Research, innovation, and strategic development of AI

The EU supports AI innovation while ensuring that that development is aligned with ethical and safety standards. The Horizon Europe programme is one of the primary EU funding mechanisms for research and development with a significant focus on AI. Its key elements include:

1. Human-centric AI: The research aims to develop AI that benefits individuals and society, ensuring alignment with ethical guidelines.
2. AI testing and experimentation facilities: These facilities allow for real-world testing of AI technologies to ensure that they meet ethical and safety requirements before large-scale deployment.
3. Public sector AI: The EU supports the use of AI in the public sector, such as healthcare and education, under conditions that ensure transparency, accountability and respect for human rights. (European Commission. 2020).

6. Cybersecurity Law

In the EU model, cybersecurity is essential in AI governance. The Cybersecurity Law, which entered into force in 2019, introduced a framework for cybersecurity certification across the EU. AI systems, especially in critical sectors such as healthcare and energy, must be cybersecurity-secure to prevent hacking, data breaches or malicious use. This helps ensure that AI systems remain safe and reliable (European Commission. 2021).

7. Coordination between Member States

The Coordinated Action Plan on AI (first introduced in 2018 and updated in 2021) is a key element of the EU's AI Governance Strategy. It encourages cooperation between EU Member States with the aim of developing a competitive AI sector while respecting common ethical and legal standards. The plan includes:

1. Investing in AI research: Encourage Member States to increase funding for AI projects.
2. AI for the public good: Promote AI applications that benefit society, in particular in healthcare, the environment, and education.
3. Training and reskilling: Help European workers acquire AI-related skills and support lifelong learning to prepare for the changing labour market (European Commission. 2021).

The main objectives of the EU AI governance model are:

1. Strengthening trust: The EU model aims to build public confidence in AI technologies by ensuring transparency, accountability and ethical compliance.
2. Reducing risks: The EU seeks to prevent harm, especially in high-risk sectors, by classifying AI systems according to their risks.
3. Supporting innovation: The EU's focus on funding innovation through Horizon Europe and AI testing facilities ensures that AI development continues to grow while upholding ethical standards.
4. Protecting fundamental rights: The EU General Data Protection Regulation and the AI Law aim to ensure that AI respects privacy, dignity, non-discrimination and democratic principles (European Commission. 2021c).

The EU's strict regulatory approach has been criticised by some stakeholders who argue that it may stifle innovation and slow down the deployment of AI technologies, especially compared to more flexible models such as those in the United States. Startups and technology companies may face significant compliance costs that could deter small businesses from entering the AI market. However, the EU sees its trust-based approach as essential for the long-term development of AI that is aligned with European values (European Commission, 2021).

The EU AI Governance Model is a pioneering framework that prioritizes the ethical development of AI, risk management, and the protection of fundamental rights. Through the AI Law, the EU General Data Protection Regulation, and its Ethical Guidelines, the EU aims to ensure that AI technologies benefit society while minimizing the risks associated with their deployment. This comprehensive and proactive approach sets global standards for AI governance and influences other countries and regions in how they develop their AI policies (European Economic and Social Committee (EESC). 2017).

CONCLUSION

As we have seen, the US AI governance model emphasizes market-driven innovation and growth with sector-specific regulations that make it a hub for AI development, but lacks focused ethical safeguards. The Chinese model is state-driven, prioritizing rapid technological development and global AI leadership while using AI for state oversight and control, raising significant ethical concerns. The Japanese model focuses on human-centered AI for social good and aligns AI with ethical principles such as fairness and human dignity, but may innovate more rapidly. The Canadian model emphasizes ethical governance and human rights, particularly in the public sector, although its impact on global AI markets is less than in other regions. Each model has unique strengths and challenges that are shaped by the region's priorities for innovation, ethics, and governance. But the EU's ethics-driven model stands out because of its strong legal framework, while the US and China prioritize rapid AI development and leadership in global markets.

Table 1. Comparison of the EU AI governance approach with other models

Component	Canada	United States	China	Japan	European Union
Governance approach	Ethics - First, human rights	Innovation-first, special section	Control from above and government control	Human-centered, social	Preventive, risk-based
Ethics and transparency	Strong focus on ethics, algorithmic impact assessment	Voluntary guidelines by the private sector	Focus on human dignity and fairness	Focus on human dignity and fairness	Strong focus, ethical guidelines
Regulations	AI in the public sector to be regulated with ethical oversight	Sector-specific fragmented regulations	State-centered, government-controlled	Transparent regulation with ethical guidelines	Comprehensive Artificial Intelligence Law
Focus on innovation	Ethical innovation, public service	High focus on innovation	Rapid innovation, government-oriented	Social and human-centered innovation	Balanced with risk management
Privacy and data protection	Privacy, focus on human rights	Lacks comprehensive privacy law	Government control over data and privacy	Privacy is protected, but innovation is slow to embrace	Strong, EU General Data Protection Regulation protects data

Ambitions of world domination	Dominance of moral governance	Economic and technological dominance	Global dominance of artificial intelligence	International cooperation on the ethics of artificial intelligence	Ethics and trustworthy artificial intelligence

Source: (Research findings, 2024)

Each model has unique strengths and challenges, shaped by regional priorities on innovation, ethics, and governance. The EU's ethics-driven model stands out for its strong legal framework, while the US and China prioritize rapid AI development and leadership in global markets.

The EU's AI governance model prioritizes ethics, risk management, and human rights through strict regulations, particularly for high-risk AI systems. The EU model serves as a global leader in promoting the ethical use of AI, but has faced criticism for stifling potential innovation. The main features of the EU's approach to AI governance are:

1. Precautionary and risk-based approach: AI systems are classified according to risk categories (unacceptable, high, limited, and minimal risk) under the proposed AI law.
2. Ethical focus: Ethical guidelines for trustworthy AI prioritize fairness, transparency, human oversight, and accountability.
3. Strict regulation of high-risk AI: High-risk AI systems must comply with strict requirements for data quality, transparency, human oversight and safety.
4. EU General Data Protection Regulation and Privacy: The General Data Protection Regulation plays a central role, focusing on user privacy, data protection and control.
5. Global leadership on AI ethics: The EU seeks to lead the way in promoting AI governance frameworks that emphasize ethics and fundamental rights.

The key strengths of the EU approach to AI governance are:

1. Protection of fundamental rights: It prioritizes the protection of human rights, privacy and the ethical use of AI.
2. Trust and accountability: High-risk AI systems are designed to be transparent and ensure trust and accountability in their deployment.

The most important challenges of the EU AI governance model are the slowdown in innovation as the EU's stringent regulatory framework potentially reduces innovation due to compliance costs and stringent legal requirements.

REFERENCES

[1] Mottaqi Dastenaei, Afshin, Karami, Ali (2024), The Digital Geopolitics Depth, Tehran: Kharazmi University Press.
[2] Alhosani, K., Alhashmi, S.M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: a review. Discov Artif Intell 4, 18 https://doi.org/10.1007/s44163-024-00111-w
[3] Almeida, P. & Santos Jr, C. & Farias, J. (2021). Artificial Intelligence Regulation: a framework for governance. Ethics and Information Technology. 23. 10.1007/s10676-021-09593-z.
[4] Attard-Frost, B. & Brandusescu, A. & Lyons, K. (2024). The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. Government Information Quarterly. 41. 101929. 10.1016/j.giq.2024.101929. Available at: https://scholarship.law.umn.edu/mjlst/vol24/iss2/4
[5] Bildt, C., Mann, E., & Vos, S. (2020). The Brussels effect: The EU's digital strategy goes global. Global Policy Watch. Retrieved 11 January, 2022, from, https://www.globalpolicywatch.com/2020/02/the-brussels-effect-the-eus-digital-strategy-goes-global/#page=1
[6] Brattberg, E., Csernatoni, R., & Rugova, V. (2020). Europe and AI: leading, lagging behind or carving its own way? Carnegie Endowment for International Peace. Retrieved 11 January, 2022, from,

https://carnegieendowment.org/2020/07/09/europe-and-ai-leading-lagging-behind-or-carving-its-own-way-pub-82236

[7] Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. Expert Systems, 41(7), e13406. https://doi.org/10.1111/exsy.13406

[8] Coeckelbergh, M. (2024). The case for global governance of AI: arguments, counter-arguments, and challenges ahead. AI & Soc (2024). https://doi.org/10.1007/s00146-024-01949-5

[9] Costa, C.J.; Aparicio, M.; Aparicio, S.; Aparicio, J.T.(2024). The Democratization of Artificial Intelligence: Theoretical Framework. Appl. Sci. 14, 8236. https://doi.org/10.3390/app14188236

[10] Danaher, J. (2022). Techno-optimism: an Analysis, an Evaluation and a Modest Defence. Philos. Technol. 35, 54 (2022). https://doi.org/10.1007/s13347-022-00550-2

[11] Data Ethics Commission. (DEC). (2019). Opinion of the Data Ethics Commission. German Federal Ministry of Justice and Consumer Protection, Berlin. Retrieved 11 January, 2022, from, https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=2

[12] De Almeida, P.G.R. (2021). dos Santos, C.D. & Farias, J.S.(2021). Artificial Intelligence Regulation: a framework for governance. Ethics Inf Technol 23, 505–525. https://doi.org/10.1007/s10676-021-09593-z

[13] Donlon, JJ. (2018). The National Artificial Intelligence Research Institutes program and its significance to a prosperous future. AI Magazine 45: 6–14. https://doi.org/10.1002/aaai.12153

[14] European Commission. (2018). Declaration. Cooperation on artificial intelligence. Retrieved 11 January, 2022, from, https://ec.europa.eu/jrc/communities/sites/jrccties/files/2018aideclarationatdigitaldaydocxpdf.pdf

[15] European Commission. (2020). White paper on artificial intelligence - A European approach to excellence and trust, COM(2020)65 final. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[16] European Commission. (2021a). Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, COM/2021a/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206#footnote10

[17] European Commission. (2021b). Commission staff working document impact assessment accompanying the proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, SWD/2021b/84 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021SC0084

[18] European Commission. (2021c). Europe fit for the digital age: commission proposes new rules and actions for excellence and trust in artificial intelligence. Retrieved 11 January, 2022, from, https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682

[19] European Commission. (2021d). Artificial intelligence - ethical and legal requirements. Retrieved 11 January, 2022, from, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements_en

[20] European Economic and Social Committee (EESC). (2017). Opinion of the European Economic and Social Committee on 'Artificial intelligence — The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society. Official Journal of the European Union C/288/1. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016IE5369&from=CS

[21] Marchant, G., & Gutierrez, C. I. (2023). Soft Law 2.0: An Agile and Effective Governance Approach for Artificial Intelligence, 24 MINN. J.L. SCI. & TECH. 375

[22] Gong, J., Qu, H., Dorwart, H. (2024). AI Governance in China: Strategies, Initiatives, and Key Considerations. https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations. Retrieved on october 12, 2024

[23] Hunkenschroer, A.L., Kriebitz, A.(2023).Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. AI Ethics 3, 199–213 https://doi.org/10.1007/s43681-022-00166-4

[24] Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nat Mach Intell 1, 389–399 https://doi.org/10.1038/s42256-019-0088-2

[25] Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. Defense & Security Analysis, 35(2), 147–169. https://doi.org/10.1080/14751798.2019.1600800
[26] Tallberg, J., & Erman, E., & Furendal, M., Geith, J., Klamberg, M., Lundgren, M., (2023). The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research, International Studies Review, Volume 25, Issue 3, September 2023, viad040, https://doi.org/10.1093/isr/viad040
[27] Jorge Ricart, R., Van Roy, V., Rossetti, F., Tangi, L. (2022). AI Watch - National strategies on Artificial Intelligence: A European perspective, 2022 edition, EUR 31083 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-52910. doi:10.2760/385851, JRC129123.
[28] Liu, Y., Zheng, W., Su, Y. (2024). Enhancing Ethical Governance of Artificial Intelligence Through Dynamic Feedback Mechanism. In: Sserwanga, I., et al. Wisdom, Well-Being, Win-Win. iConference 2024. Lecture Notes in Computer Science, vol 14598. Springer, Cham. https://doi.org/10.1007/978-3-031-57867-0_8
[29] McLean, S., Read, G. J. M., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2021). The risks associated with Artificial General Intelligence: A systematic review. Journal of Experimental & Theoretical Artificial Intelligence, 35(5), 649–663. https://doi.org/10.1080/0952813X.2021.1964003
[30] Munthe, C. (2020). Precautionary Principle (Version 2). 10.1002/9781444367072.wbiee550.pub2.
[31] Novelli, C., Casolari, F., Rotolo, A. et al. (2024). AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act. DISO 3, 13 https://doi.org/10.1007/s44206-024-00095-1
[32] Pflanzer, M., Dubljević, V., Bauer, W.A. et al. (2023). Embedding AI in society: ethics, policy, governance, and impacts. AI & Soc 38, 1267–1271 https://doi.org/10.1007/s00146-023-01704-2
[33] Roberts, H., Cowls, J., Morley, J. et al. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. AI & Soc 36, 59–77 https://doi.org/10.1007/s00146-020-00992-2
[34] Roberts, H. and Babuta, A. and Morley, J. and Thomas, C. and Taddeo, M. and Floridi, L., (2023). Artificial Intelligence Regulation in the United Kingdom: A Path to Good Governance and Global Leadership? Internet Policy Review, Available at SSRN: https://ssrn.com/abstract=4209504 or http://dx.doi.org/10.2139/ssrn.4209504
[35] Science, Innovation and Technology Committee, (2024). Governance of artificial intelligence (AI). https://publications.parliament.uk/pa/cm5804/cmselect/cmsctech/38/report.html, Retrieved on october 12
[36] van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. AI Ethics 1, 213–218. https://doi.org/10.1007/s43681-021-00043-6
[37] Veale, M. & Matus, K. & Gorwa, R. (2023). AI and Global Governance: Modalities, Rationales, Tensions. Annual Review of Law and Social Science. 19. 10.1146/annurev-lawsocsci-020223-040749.
[38] Zaidan, E., Ibrahim, I.A.(2024). AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective. Humanit Soc Sci Commun 11, 1121 https://doi.org/10.1057/s41599-024-03560-x
[39] Züger, T., Asghari, H. (2023). AI for the public. How public interest theory shifts the discourse on AI. AI & Soc 38, 815–828 https://doi.org/10.1007/s00146-022-01480-5