DOI 10.18276/978-83-8419-028-9-37

28th European Conference on Artificial Intelligence (ECAI 2025)

ETHICS AND GOVERNANCE IN AI MANAGEMENT

Aleksandra Klich[a], Bartosz Brożyński[b], Yuliia Khvatsik[c], Marcin Białecki[d]

[a] *University of Szczecin, Institute of Legal Sciences, Poland*
[b] *University of Szczecin, Institute of Management, Poland*
[c] *Jönköping University, Business School, Sweden*
[d] *Cardinal Stefan Wyszyński University, Institute of Legal Sciences, Poland*

ABSTRACT

Purpose: *The publication aims to analyse ethics and governance in artificial intelligence (AI) management within organisations. It examines how ethical principles and governance mechanisms facilitate the responsible use of AI, taking into account both legal obligations and social expectations.*

Need for the study: *The study addresses the growing significance of AI in various sectors and the challenges it presents to society, the economy, and individual rights.*

Methodology: *An interdisciplinary approach is used, combining a literature review, legal analysis, and case studies. The study integrates legal, ethical, and technological perspectives to provide a comprehensive view of AI governance.*

Findings: *Three governance models are examined: regulatory, self-regulatory, and hybrid. The regulatory model offers legal certainty but limited adaptability. The self-regulatory model allows innovation and flexibility but may lack consistency in ethical norms. The hybrid model, which combines both approaches, is the most effective, striking a balance between structure and flexibility, and sharing responsibility.*

Practical Implications: *The article discusses the implementation of effective oversight mechanisms tailored to industry-specific needs and technological maturity. It emphasizes the need for a dynamic regulatory framework that adapts to technological advancements, while promoting self-regulation and private sector accountability. A balanced governance model—combining formal legal controls with voluntary industry standards—is essential for promoting safety, transparency, and ethical compliance in AI use.*

Keywords: Artificial Intelligence; Governance; Ethical; Management; Legal Rules

Jel codes: K23, Regulated Industries and Administrative Law; K30, General; K49, Other

1. INTRODUCTION

Artificial intelligence (AI) is becoming an increasingly common tool in organisational management, supporting decision-making, data analysis, process automation, and resource optimisation. It is becoming a key element of modern management, transforming decision-making processes, optimising organisational activities, and redefining corporate structures. AI systems are

Peer-review under responsibility of the scientific committee of 28th European Conference on Artificial Intelligence (ECAI 2025) and Intelligent Management Workshop.

used in finance, healthcare, public administration, and human resource management, which opens up a wide range of possibilities, but at the same time raises serious ethical and legal challenges. AI encompasses a set of technological capabilities that provide new resources, opportunities, and applications for organisations and society (Berente et al., 2021). However, significant ethical and regulatory challenges arise with AI systems' increasing autonomy and influence. However, the rapid spread of artificial intelligence systems has given rise to risks associated with algorithmic opacity, unethical behaviour, and unintended consequences (Kroll et al., 2017). These concerns also relate to the potential of generative artificial intelligence to replace humans in the workforce, including issues such as technostress, bias in artificial intelligence algorithms, lack of transparency, and potential misuse of artificial intelligence technologies (Korzyński et al., 2024). The key issue, therefore, is to create supervisory mechanisms that will ensure the responsible implementation of AI, combining innovation with transparency and efficiency with responsibility. This involves establishing ethical, technical, social, and legal requirements to ensure that algorithms and data comply with standards, while also establishing ethical principles, acceptable practices, and regulations (LaBrie et al., 2019).

On the one hand, AI contributes to increased efficiency and precision in management by eliminating subjective human errors. On the other hand, its functioning can lead to a lack of transparency in decisions, algorithmic discrimination, violations of employee rights, and threats to privacy protection. The increase in AI autonomy raises key questions regarding legal responsibility for erroneous decisions made by algorithms. These issues require appropriate regulations and oversight mechanisms to control the technology and ensure its legality. The literature on the subject expresses the view that the expected increase in digital reliability and security incidents will exceed 1,000 cases per month by 2030. Incidents related to transparency, explainability, and accountability will also increase steadily, emphasising the need for a governance framework that prioritises these principles (Cebrian et al., 2025).

Currently, AI systems are subject to various regulations, depending on the jurisdiction and sector in which they are used. In the European Union, the AI Act Regulation plays a key role in artificial intelligence regarding risk and ensuring its compliance with fundamental rights. Other critical legal acts are: a) General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) – imposing an obligation to protect personal data processed by AI; b) The Directive on Liability for Defective Products (Directive (EU) 2024/2853), which may apply to erroneous decisions made by AI that lead to damages; c) European Charter of Digital Rights (European Declaration on Digital Rights and Principles for the Digital Decade), which defines the rights of users about digital technologies, including AI. Work is underway at the global level to harmonise regulations. The Organisation for Economic Co-operation and Development (OECD) and the United Nations (UN) are developing guidelines for the ethical use of AI. In the USA and China, regulatory approaches differ - the USA focuses on self-regulation of the sector, while China is introducing strict control over AI, especially in the context of its use in administration and national security.

The publication aims to analyse the challenges of AI governance, focusing on legal and regulatory aspects. In particular, the article focuses on identifying key legal and ethical issues related to AI, such as legal liability and compliance with data protection and human rights regulations. The authors analyse current legal regulations concerning AI, including the EU AI Act, GDPR, and national rules in selected countries. The aim is also to assess the effectiveness and impact of existing supervisory mechanisms on organisations' management practice. It also aims to formulate recommendations for effective legal solutions and governance models that can ensure the responsible implementation of AI in management. The basic research hypothesis is that current legal regulations concerning AI are insufficient, leading to the risk of abuse and a lack of responsibility for algorithm decisions. At the same time, in the authors' opinion, the lack of transparency of AI algorithms hinders effective governance and enforcement of regulations, which can lead to violations of human rights and business ethics. In contrast, implementing comprehensive AI audit and governance mechanisms can improve the quality of management and limit the risk of discrimination and abuse. Undoubtedly, EU regulations, such as the AI Act, can become a global model for AI regulation, influencing the shaping of technology policy worldwide.

AI governance is one of the key challenges of the 21st century. The lack of effective regulation can lead to serious legal and reputational consequences for organisations, such as sanctions for GDPR violations, lawsuits regarding algorithmic discrimination, or liability for AI decisions resulting in financial losses. Regulations such as the AI Act attempt to balance innovation and protecting

fundamental rights. Still, their effectiveness depends on implementing appropriate enforcement and auditing mechanisms. Organisations must therefore invest not only in AI development, but also in legal and ethical compliance systems to ensure the responsible management of this technology. The rest of this paper will analyse specific cases of AI use in management, assess the effectiveness of regulation, and present solutions for improving AI governance.

2. LITERATURE REVIEW

The issue of using artificial intelligence to improve decision-making processes and increase productivity is increasingly being addressed in the literature. Research focuses on using AI to identify patterns, analyse scenarios based on forecasts, learn from historical data, and recommend actions to optimise performance (Akerkar, 2019). The growing number of publications in this area reflects the increasing interest in the theoretical aspects of this technology and its practical applications in various sectors. The ever-increasing number of publications aligns with the growing number of publications analysing how artificial intelligence affects business models and corporate strategies. The academic study aims to investigate the role of artificial intelligence in transforming business models, emphasising the interaction between technological innovation and business strategy. The authors emphasise the long-term consequences of technological innovation for the economy and market competitiveness, and in the area of necessary legal changes (Oluwatoyin et al., 2023).

The topicality of artificial intelligence in various sectors is also determined by the fact that this issue is increasingly becoming the subject of scientific publications. This is in response to practice, as this field is dynamically developing. J. Laine, M. Minkkinen, and M. Mantymaki (Laine et al., 2024) review the literature by analysing and synthesising various approaches to ethical principles in the context of AI auditing and present relevant information for key beneficiaries of this process. Thanks to the publication, it is possible to identify the importance of such features as fairness, transparency, non-maleficence, responsibility, privacy, trust, beneficence, and freedom and autonomy in AI systems. In the authors' opinion, it is possible to distinguish three basic types of knowledge, i.e., guidelines, methods, tools, and frameworks for action, as well as awareness and increased capacity for action. Publications that address issues related to the use of AI contribute to the literature review by combining an analysis of the impact of artificial intelligence on education and the service market in the context of digitalisation and dynamic technological progress. Publications such as D. Badulescu et al. (Badulescu et al., 2025) provide theoretical justification and empirical evidence for adapting educational competencies to the requirements of the AI-driven labour market.

In literature, there is a trend towards analysing AI management methods from the perspective of specific sectors (e.g., medical, banking, agricultural, etc.). This is because each industry has unique needs, challenges, and regulations. This emphasises the fact that AI does not work in isolation. Its application in healthcare differs from its use in agriculture or law. For example, R. Dara, F. S. M. Hazrati, and J. Kaur (Dara et al., 2022) highlight the growing role of AI in digital agriculture, emphasising both its benefits and the ethical challenges it poses. The authors emphasise that AI can improve the efficiency of agricultural operations and support decision-making. On the other hand, they emphasise that its improper implementation can lead to negative consequences, such as violation of farmers' privacy, threat to animal welfare, or lack of responsibility for the effects of AI systems. Specific considerations complement the general considerations regarding AI's ethical aspects, transparency principles, and personal data protection.

3. METHODOLOGY

Studying ethics and governance in AI management requires a multifaceted methodological approach, including legal perspectives and ethical and managerial aspects. During the work on this publication, a multidimensional research approach was applied, combining the analysis of legal regulations, scientific achievements, and practical challenges related to the implementation of AI systems. The following research methods were used. Firstly, the dogmatic-legal method analyzes current legal norms, their interpretation, and application in practice. It focuses on the

systematisation of regulations and their interpretation. This method was used to analyse existing legal regulations, the understanding of standards concerning ethical and legal AI governance (e.g., the principle of transparency, algorithmic accountability, the right to explain AI decisions), as well as to examine the relationship between regulations and AI management practice in regulated sectors. Analytical methods were also applied to logically explore legal and ethical concepts, principles, and structures. Applying this method, the authors thoroughly reviewed the literature on the subject. This method was used to analyse ethical concepts applied in AI, verify the effectiveness of AI governance supervision, identify the risks associated with using AI systems, and analyse the risks associated with the lack of appropriate governance models.

The comparative law method was applied to a basic extent due to the lack of final regulations at the level of national legislators. In this respect, its use refers to comparing bare legal acts at the international, EU, and, in some cases, national level. This analysis aimed to identify similarities, differences, best practices, and the consistency of different legal acts regulating identical issues. Using this method, the authors compared EU regulations on AI with solutions used in the USA, China, or the UK, and analysed AI governance models. The result of this analysis was the identification of three models of AI governance, i.e., regulatory, self-regulatory, and hybrid, with an indication of their effectiveness in the areas studied. To a minimal extent, an empirical method was applied by observing real challenges related to the regulation and implementation of AI, which was possible thanks to the Authors' professional experience in the legal and management fields. The experiences of practitioners and cases of AI application in various sectors, such as healthcare, the financial industry, and the judiciary, were analysed.

The applied research approach allowed for a comprehensive assessment of the ethical and regulatory challenges in managing artificial intelligence. The integration of legal analysis, literature review, and observation of practical experiences made it possible to formulate conclusions relevant to both the academic community and entities involved in implementing and regulating AI systems.

4. AI MANAGEMENT GOVERNANCE

The dynamic development of AI is associated with both enormous innovative potential and risks to individual rights, public safety, and economic stability. In response to these challenges, various AI governance models have been developed to ensure that AI systems comply with legal, ethical, and technological standards. For this study, governance models have been distinguished based on the degree of state interference and AI implementers' autonomy level. As a result, three basic governance models have been identified: regulatory, self-regulatory, and hybrid. The rationale for adopting the above classification of AI governance model is primarily based on analyzing the degree of state interference (from centralised in the case of the regulatory model, through strong in the self-regulatory model, to mixed in the hybrid model). The level of autonomy of the entities implementing AI is also necessary, as the degree of freedom left to entrepreneurs and organisations in shaping the principles of responsible implementation of AI-related mechanisms is a crucial factor. In addition, different governance models have been noted in practice, depending on a given industry's risk and regulatory needs. As a result, the classification adopted allows for an analysis of the advantages and disadvantages of each model, their impact on innovation, security, and the development of AI.

4.1. Regulatory governance

The first of these, regulatory oversight, is based on strict legal regulations and enforcement mechanisms at the international, European, and national levels. This oversight model is based de facto on rules set at a higher level, obliging national legislators to implement them through state institutions and market governance bodies. This model can be described as centralised or direct. Under this model, the state legislates the rules for designing, implementing, and using AI, imposing obligations on entities regarding compliance, reporting, and auditing. An example of this approach is the EU AI Act, which classifies AI systems according to risk level and imposes different legal requirements. It is worth emphasising that despite the significant development of artificial intelligence and the large increase in its users worldwide, there was still no comprehensive legislation governing the situation of AI and its users until mid-2024. This situation changed on 1 August 2024 when the AI Act, published on 12 July 2024, came into force. It is the world's first

act to regulate artificial intelligence in such a comprehensive manner. The document in the form of a Regulation regulates safety, transparency, and compliance, the observance of which is to ensure the development and application of AI following the values of the European community. Due to the broad subject matter regulated in the AI Act, most provisions will come into force 24 months after its announcement. However, individual regulations will gradually come into force after 6, 12, or even 36 months. For example, rules banning the use of dangerous systems throughout the EU came into force in February 2025, and AI governance regulations will come into force in August 2025. The impetus for adopting the above act was the desire to harmonise rules on digital services in EU member states, thus avoiding fragmentation of the internal market. This harmonisation covers intermediary services in the internal market. It aims to ensure a safe, predictable, and trusted online environment, to counter the dissemination of illegal content online, and to address the social threats that the dissemination of disinformation or other content may pose (recital 9 of the AI Act). The Digital Services Act aims to build a safe and fair digital environment (Urbanowicz, 2025). Among intermediary services, the EU legislator in Article 3(g) of the AI Act distinguishes the service of: a) 'mere conduit' (transmission in a telecommunications network of information provided by a recipient of the service, or the provision of access to a telecommunications network); b) 'caching' (transmission in a telecommunications network of information provided by a recipient of the service, including the automatic, intermediate and temporary storage of that information, carried out solely to facilitate the subsequent transmission of the information upon request by other recipients); c) 'hosting' (storage on the information provided by a recipient of the service and at their request).

The wording of recital 153 of the AI Act is crucial from the perspective of the regulatory model, in which the EU legislator stipulates that the Member States play a key role in the application and enforcement of the regulation. In each national legal system, at least one notifying authority and at least one market surveillance authority should be designated as the national competent authority to supervise the application and enforcement of the AI Act. The EU legislator authorises Member States to establish any public entity to perform the tasks of national competent authorities, which, according to Article 3(38) of the AI Act, are notifying authorities or market surveillance authorities. By the regulation mentioned above, concerning AI systems put into use or used by EU institutions, bodies, and organisational units, the references to national competent authorities and market surveillance authorities contained in the AI Act should be read as references to the European Data Protection Supervisor. It is worth emphasising here that the AI Act and other EU data protection regulations (in particular the GDPR (Regulation (EU) 2016/679, 2016), EUDPR (Regulation (EU) 2018/1725, 2018) i LED (Directive (EU) 2016/680, 2016), and the e-Privacy Directive (Directive 2002/58/EC, 2002) must, in principle, be treated (and interpreted consistently) as complementary and mutually reinforcing instruments. The EU data protection regulations apply to processing personal data throughout the entire life cycle of AI systems, as explicitly stated in Article 2(7) and recitals 9 and 10 of the AI Act.

According to recital 10 of the AI Act, this regulation does not affect or modify the existing European Union regulations on protecting personal data, including the competences of independent governance bodies responsible for monitoring compliance with these regulations. It also does not change the obligations of providers and users of artificial intelligence systems who act as controllers or processors under national or EU law on protecting personal data if AI systems' design, development, or use involves processing such data. In addition, the AI Act does not limit the rights and safeguards of data subjects under EU law, including those related to decisions made solely by automated means, including profiling. According to Article 2(1) of the AI Act, this regulation applies to intermediary services offered to service recipients established or located in the EU, regardless of the place of establishment of the providers of these intermediary services. The EU legislator has not defined a closed list of purposes that a digital service must fulfil from the user's perspective to be considered an intermediary service. For this reason, it is worth emphasising at the outset that, in principle, any digital service that enables the transmission and storage of data will be considered an intermediary service, provided that it meets the criteria defined in the DSA (Urbanowicz, 2025). An intermediary service provider is any entrepreneur who offers these services on the European market. This means that the provider's activity consists of directing the services they provide to people in the EU, regardless of the country in which the provider operates.

Another characteristic of the regulatory model is that the EU legislator recommends that Member States designate, following national law, one or more authorities and entrust them with supervising the

application and enforcement of the AI Act. The recommendations of the AI Act, as expressed in recital 109, give EU countries flexibility in the extent to which they can assign tasks and competences to existing bodies in specific sectors. Consequently, various legal forms of governance bodies are permissible, meaning they can function within already existing administrative structures. This sectoral approach allows for the assignment of specific competences to bodies already operating in the relevant areas, e.g., electronic communications authorities if the regulation concerns telecommunications, media regulatory authorities if it concerns the governance of media content, or consumer protection authorities if the regulation relates to the protection of consumer rights. It can therefore be assumed that the EU legislator, in creating a general framework for governance in AI management, respects the diversity of constitutional and administrative structures in the Member States, leaving them free to organise governance. As a result, this can lead to different governance models in individual countries, e.g., centralised, where one body supervises the entire regulation, or decentralised, where competences are distributed among several institutions. However, coordination between the various authorities is crucial to avoid confusion over competences and the dispersion of responsibilities. In practice, this may require the creation of a framework for cooperation and information exchange between authorities, e.g., through the conclusion of inter-agency agreements, or the creation of joint supervisory and sanctioning procedures. Thus, the EU legislator focuses on flexibility and adaptation to national realities, while requiring effective governance. The key challenges are ensuring consistency in enforcement and effective coordination between the various authorities, especially in countries that divide competencies.

It is also worth noting that in recital 113 of the AI Act, the EU legislator gives EU countries the option of designating an existing national authority to act as a Digital Services Coordinator or to perform specific tasks to oversee the application and enforcement of the AI Act. This is conditional on such a body meeting the requirements set out in the regulation, including, in particular, its independence. This may require an adjustment of the body's competences, e.g., by strengthening its operational and financial autonomy. The regulation allows Member States to combine different regulatory functions in a single body, if compatible with EU law. For example, the governance functions over digital platforms can be assigned to a body that governs media, telecommunications, or consumer protection. However, as mentioned above, the recital clearly states that it must not weaken their independence if the state decides to reorganise the governance bodies. The EU legislator specifies in this respect that the chair or board members of a collegial body cannot be dismissed before the end of their term of office solely because of an administrative reform if there are no guarantees of their continued independence. This regulation aims to protect against a situation in which the government could interfere with the composition of the bodies to increase political influence on their activities. When analysing the above recommendations and principles created by the EU legislator, it is reasonable to assume that Member States are given flexibility while maintaining the independence of governance bodies. This is particularly important in digital services, where effective governance of online platforms requires autonomy from political pressure and market interests. Member States may therefore reform their governance structures, but they must do so in a way that does not undermine regulatory effectiveness or public trust. This provision allows flexibility in the governance of digital services but sets precise requirements for the independence of governance authorities. Institutional reforms must not be used to remove inconvenient regulators or weaken law enforcement.

Consequently, it should be emphasised that Member States are required to designate or establish three types of authorities in the implementation of the EU AI Act, i.e.:

a) Market Surveillance Authority (MSA), whose task will be to implement the measures provided for in Regulation (EU) 2019/1020 on market surveillance and product compliance; this institution is based on the already existing and well-established concept of market surveillance authorities in EU law, and its main task will be to ensure that only products that meet EU legal requirements are on the EU market (Art. 3 No. 26 AI Act);
b) Notifying Authority (NA), i.e. the national authority whose task will be to establish and implement a procedure for the assessment, designation and notification of conformity assessment bodies, as well as for their governance (Article 3(19) and Article 28(1) AI Act); in this group of entities, ‘Conformity assessment bodies’ should also be distinguished. Entities performing third-party conformity assessment activities, including testing, certification, and inspection (Article 3(21) AI Act). According to Article 3(48) of the AI Act, the notifying and market surveillance authorities are jointly referred to as National

Competent Authorities (NCAs). Consequently, they must function independently, impartially, and without bias, have adequate technical, financial, and human resources and infrastructure to effectively perform their tasks by the Artificial Intelligence Act, as confirmed by the wording of Article 70(1) and (3) of the AI Act.

c) National Public Authorities (NPAs), which will be responsible for enforcing the obligation to respect fundamental rights in the Member States in the context of high-risk AI systems; they should have the power to request and access any documentation drawn up or stored under the AI Act, if this is necessary for the effective fulfilment of their mandate within their jurisdiction (Art. 77 (2) AI Act).

Transferring the above considerations to the practical aspects of implementing the AI Act in member states, it is worth noting that the European Commission established the European Artificial Intelligence Authority in February 2024. This authority supervises the enforcement and implementation of the act on artificial intelligence in EU member states. Some EU countries have already taken steps towards establishing governance authorities responsible for their implementation. According to the AI Act, Member States are required to designate market surveillance authorities at the national level by 2 August 2025. Legislators in the Member States are working on creating appropriate mechanisms and governance bodies to supervise the application and implementation of the AI Act. Each Member State must designate the appropriate authorities and publish a list.

In September 2023, Spain, by Royal Decree No. 729/2023 (Real Decreto 729/2023, 2023), established an AI governance agency (esp. *Agencia Española de Supervisión de Inteligencia Artificial,* AESIA) based in A Coruña. AESIA is a specialised body responsible for governance, advice, awareness raising, and training related to the development and use of artificial intelligence at the national level. Its tasks include ensuring compliance with national and EU regulations on AI, coordinating the implementation and application of the regulation's provisions, and cooperating with the EU AI Office and market surveillance authorities from other Member States. In addition, AESIA promotes AI innovation in the public and private sectors by creating regulatory sandboxes and developing best practices. In other EU member states, discussions are taking place on the designation of appropriate governance authorities for the AI Act. The position of the European Data Protection Board (Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, 2024) is worth mentioning, according to which data protection authorities should play an essential role in enforcing the AI Act. In this context, these bodies are recommended to act as market regulators, especially for high-risk AI systems used in law enforcement, border management, and the judiciary (Statement 3/2024). The justification for this position is that the data protection authorities, in connection with the EU General Data Protection Regulation, have experience and knowledge of the impact of AI on fundamental rights. As of today, 23 EU countries, i.e. Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Portugal, Romania, Slovakia, Slovenia, Spain and Sweden have appointed the relevant authorities and published a list of tchem (Artificial intelligence act management and enforcement). When analysing the overall status of National Authorities, it should be emphasised that as of 8 November 2024, the unclear status of authorities concerned 16 MSAs and NCAs and 12 NPAs, partial clarity was identified for 10 MSAs and no NCAs, and clarity concerns 1 MSA and 15 NCAs (Overview of all AI Act National Implementation Plans, 2024). This data clearly shows the ongoing work of national legislators. The Polish legislator is also developing rules for implementing the AI Act. During the pre-consultation organised by the Ministry of Digital Affairs, most participants favored creating a new market governance body instead of transferring these powers to existing institutions, such as the President of the Office for Personal Data Protection. In response to these demands, the Ministry of Digital Affairs announced the creation of an AI governance committee, which is reflected in the draft law on artificial intelligence systems (Draft Polish law on artificial intelligence systems, 2024, October 17, No. UC71, 2024). The Commission for the Development and Security of Artificial Intelligence (KRiBSI) is intended to be the market governance authority for artificial intelligence systems within the meaning of Section 70 (1) of the AI Act. It will act as a single point of contact. According to the will of the initiator, KRiBSI is to cooperate with several existing authorities and ministries. In addition to its governance activities, the Commission's role is primarily to strengthen the artificial intelligence industry and act as an expert entity cooperating with other public institutions in AI.

In conclusion, it seems reasonable to assume that the regulatory model for AI management governance created by the EU legislator is quite demanding, while at the same time giving Member States freedom to shape the structure of entities equipped with governance competences. The AI Act defines a general legal framework, but leaves room for national legislators to clarify the regulations. There is no doubt that, despite the ongoing technological process, many EU countries do not yet have sufficient technical and human resources to analyse and supervise AI systems effectively. National legislators in EU countries face numerous challenges in implementing the AI Act, including legal, organisational, and technological aspects. A lack of adequate human and financial resources can slow the process, and differences in approach between EU countries can lead to market fragmentation. However, it is clear that in Poland and other EU countries, the implementation of the AI Act is in progress, with varying degrees of progress and challenges. However, there is no doubt that each Member State faces the challenge of implementing the AI Act in a way that, on the one hand, provides adequate protection against AI-related risks and, on the other hand, prevents the inhibition of innovation.

4.2. Self-regulatory governance

The second AI governance model distinguished for this study is self-regulatory governance. It includes all internal standards that establish artificial intelligence standards in the activities of entities using AI systems. Legislative measures taken at the EU level have therefore given private sector companies an incentive to get involved in developing and implementing artificial intelligence (Taebi, 2019), and international efforts are currently being made to achieve this goal. It must be admitted that China was the first country to regulate AI. However, due to the two divergent objectives of China and the European Union, it is considered that the AI Act is the world's first act regulating issues related to artificial intelligence. China has chosen to strengthen state control over the content provided to citizens, while the EU has set itself the goal of minimising the damage caused by the use of AI by European citizens (Maciążka, 2025). Other non-EU countries also recognise the need to regulate AI. The UK is one such country. However, in this case, it was decided that the regulations should support the development of artificial intelligence rather than restrict it. Therefore, the rules adopted are selective, more flexible, and certainly not comprehensive. In the USA, there are currently no federal regulations on AI. However, there are regulations in individual states; bills in this context have been passed in 45 states. The legislative situation regarding artificial intelligence is also influenced by political decisions leading to the repeal of the executive order on the safe and reliable development and use of artificial intelligence. President Donald Trump has already announced that he will begin work on AI regulations. Still, they will be aimed primarily at strengthening the American AI industry, which may mean he will not continue with the concept proposed by Joe Biden.

Significantly, since 2 February 2024, suppliers and AI system users have been obliged to ensure that their employees have the appropriate AI knowledge and skills necessary to perform their tasks. This obligation covers the area of AI competence, so-called AI Literacy (Article 4 AI Act). According to the regulation mentioned above, providers and users of artificial intelligence systems should ensure adequate knowledge and skills in AI among their personnel and persons responsible for the operation and use of these systems. The level of required competences should take into account: the level of technical knowledge, professional experience, level of education and training, the specifics of the context in which AI will be used, as well as the characteristics of individuals or groups to whom AI systems will be applied. Thus, the EU legislator emphasises the need to ensure adequate competence among persons and entities responsible for implementing artificial intelligence systems. This is crucial from the perspective of minimising the risk of errors, as a lack of technical knowledge can lead to the misuse of AI systems. This, in turn, can result in violations of personal rights, discrimination, or incorrect algorithm decisions. It is also essential to ensure the safe and ethical use of AI. Proper personnel training helps to consciously implement and supervise AI systems by legal and ethical standards. It is worth noting here that doubts about industry self-regulation also characterise the current debate on the ethics of AI. Dealing with the moral challenges of artificial intelligence has become a priority for legislators worldwide. This issue is analysed later in this paper. Ensuring that those who manage and use AI have the right skills is essential to adapt skills to the context and audience. EU legislators emphasise that knowledge requirements should consider the specific nature of the situation in which AI is used; for example, different skills will be required in healthcare than in the financial

sector. The obligation to ensure employee competence means that organisations cannot use a lack of knowledge as an excuse for AI errors or malfunctions, which positively strengthens the responsibility of both AI providers and users. The self-regulatory model means that those who use and manage AI must determine what educational and training measures are necessary for implementation and what promotes self-regulatory governance mechanisms in AI management. The justification for creating internal regulations for AI management in a broad sense is therefore the pursuit of responsible and ethical use of this technology.

Regarding the text of the AI Act, recital 88 of the aforementioned legal act is also of significant importance, in which the EU legislator obliges providers of huge online platforms and substantial online search engines to exercise due diligence in the measures taken for testing. It is recommended that they adapt their algorithmic systems, particularly their recommendation systems, if necessary. The aim of implementing self-regulatory solutions is to ensure that AI systems are developed and implemented ethically, safely, and in accordance with applicable laws and social standards. As indicated, the AI Act has broad application for companies in various industries, not only those related to IT or artificial intelligence. The new regulations cover suppliers, distributors, and importers from all EU member states. This, in turn, generates several obligations for them. Suppliers of AI systems, especially those considered high-risk, must carry out a risk assessment, submit technical documentation, and guarantee the system's highest level of security and transparency. In addition, they must systematically audit compliance and adapt subsequent technologies to the requirements specified in the AI Act. This means these entities must develop a clear and systematic approach to regulatory compliance, affecting their internal structure and day-to-day operations.

This means that processes must be implemented to identify, assess, and minimise the risks associated with artificial intelligence. This will be reflected in the creation of risk management procedures. Each new version of an AI system should be evaluated for potential risks (e.g., discrimination, unforeseen decisions). Artificial intelligence system providers should also monitor whether AI systems cause unintended consequences (e.g., misjudgment of customers in credit systems). Within this area of self-regulation, it seems reasonable to establish and adopt procedures to be followed if irregularities are detected, including the possibility of immediate intervention in the event of AI malfunctioning. It is also important to create methods to control the technical documentation of AI systems, which should undoubtedly be created and updated. This documentation should include descriptions of the system's operation and algorithms, testing and validation methods, and measures to ensure compliance with regulations, including the AI Act. Simply creating internal regulations without regularly reviewing them will be pointless. For this reason, it is essential to regulate the conduct of compliance audits and AI quality control. In practice, AI compliance teams should be separated in the organisational structure of AI system providers responsible for monitoring and controlling processes. In this context, these entities may be required to adopt external certifications confirming the compliance of their systems with EU regulations.

Distributors and importers, on the other hand, are responsible for verifying that the AI systems they place on the market comply with the legal regulations in the AI Act. In the event of violations, they may bear similar liability to suppliers. As for users, their situation is also regulated in the AI Act. Their responsibilities include monitoring AI systems, reporting problems, and confirming that the technologies provided are being used for their intended purpose and in accordance with the supplier's instructions (Naron-Grochalska, 2024). In light of the above, it is possible to assume that the AI Act places great emphasis on self-regulation by obliging entities to create internal regulations and procedures to ensure compliance with new legal requirements and the ethical and responsible use of artificial intelligence. However, it is still necessary to continue researching the moral management of artificial intelligence and how companies can create value that benefits both business and society (Naron-Grochalska, 2024).

The self-regulatory governance model assumes that AI system providers are responsible for developing and implementing internal regulatory compliance oversight systems, including risk assessments, risk mitigation strategies, and specific countermeasures. The EU legislator's call for creating codes of conduct is intended to facilitate the independent monitoring of and response to potential law violations. It can therefore be assumed that internal supervision is de facto based on the organisation's independent commitment to creating and implementing ethical and legal standards. Suppliers are obliged to ensure transparency of their activities and to provide mechanisms to assess the

effectiveness of their governance systems. It is worth emphasising, however, that the self-regulatory governance model, although more flexible than regulatory governance, carries the risk of inconsistent application of standards and difficulties in ensuring complete transparency and accountability.

4.3. Hybrid governance

The third and final supervision model is a hybrid model, which combines public and private elements. In this case, we are talking about recommendations and guidelines constituting so-called soft law, addressed primarily to AI system providers regarding desirable solutions in internally created procedures and standards. Soft law guidelines are voluntary measures regulating the ethical development and implementation of artificial intelligence (Bietti, 2020). Cooperation between public governance and self-regulatory systems is crucial in the hybrid governance model. This applies to both the exchange of information, the assessment of risks, and the conduct of audits. In the former case, the legislator will define the rules for managing specific areas, and suppliers must provide information on their activities. In the case of risk assessment and auditing, both external (regulatory) and internal (self-regulatory) governance should cooperate in conducting risk assessments and audits to identify potential sector-specific risks (e.g., in the case of advertising, this concerns manipulation and disinformation). In this model, the EU or national legislator suggests the necessary measures, but the final decision-maker in this context is the entity managing the AI system. It is up to these entities to decide whether the recommended solutions will be introduced into the supervision. Recital 21 of the AI Act emphasises that the European Commission and the Member States, cooperating with relevant stakeholders, should facilitate the development of voluntary codes of conduct to improve AI competence among those involved in developing, operating, and using AI. This is a direct manifestation of the hybrid model of AI governance and AI management.

This requirement is expressed in recital 117 of the AI Act, as the EU legislator stipulates that codes of practice should be one of the main tools for properly fulfilling the obligations under the AI Act for providers of general-purpose AI models. Suppliers should be able to rely on codes of practice to demonstrate compliance with these obligations. The EU legislator indicates that codes of practice are the primary tool enabling providers of general-purpose AI models to demonstrate compliance with the obligations under the AI Act. This mechanism is crucial for self-regulation because it allows AI providers to develop compliance rules. Looking at the codes of good practice in personal data protection in individual sectors created in connection with the GDPR, it is possible that the codes of practice are not rigid regulations imposed by law. They are industry standards that can be developed by AI providers themselves or sector organisations. Due to the diversity of AI systems and the sectors for which they are implemented, codes of practice will enable a differentiated approach depending on the type of AI model. This means that AI system providers can adapt risk assessment, testing, transparency, and security procedures to the specifics of their systems.

Despite the wide range of freedom in the actions taken by AI system providers, the fundamental rule that codes of good practice must be in line with the general objectives of the AI Act should not be overlooked. Importantly, self-regulatory governance can occur under the supervision of the European Commission and the AI Office, as the AI Act provides in recital 117 that the EC can endorse codes of practice, giving them general validity in the Union. However, if the codes are not created or prove insufficient, the EC can introduce implementing acts, which means the possibility of regulatory interference in the event of ineffective self-regulation. However, if the code of practice is approved, its application implies a presumption of compliance with the AI Act, which limits the risk of sanctions and regulatory controls. However, AI suppliers can also use alternative means of demonstrating compliance, giving them freedom of choice within the self-regulation framework.

One example of an area where such recommendations would be justified is the advertising sector. As indicated in recital 95 of the Digital Service Act (DSA) (Regulation (EU) 2022/2065, 2022), the use of AI systems in advertising, especially in the case of extensive online platforms (e.g. Google, Facebook) and search engines, is based on advanced algorithms that analyse vast amounts of user data to target advertising. The aim is to tailor advertising content to individual user preferences, often based on their previous interactions with the platform, online behaviour, and demographic data. These systems create the risk of information manipulation, privacy violations, or insufficient oversight. Advertisements can be used to spread misinformation, which can negatively affect public health, security, or civil discourse.

In turn, using users' data without complete transparency about what data is collected and how it is used can lead to a breach of privacy. In Article 22(3) of the GDPR, the EU legislator establishes that every user should have the right to obtain explanations and challenge AI decisions. Despite the use of advanced technology, these systems often operate in a problematic way to monitor, which minimizes the possibility of ensuring full accountability for their operation. For this reason, the necessity of public supervision over these systems is emphasised. The measures provided for by the EU legislator include advertising repositories. It is assumed that extensive online platforms and search engines should provide access to advertising repositories to facilitate public supervision. These repositories are intended to contain the content of advertisements (name of the product, service, brand), information about the advertisers and entities that paid for the ad, the targeting criteria of the advertisements, and data about the audience, including details about the targeting of vulnerable groups (e.g., minors). The repositories aim to increase transparency and enable regulators and the public to monitor and investigate the risks associated with the distribution of advertisements, including disinformation and manipulation. The hybrid governance model is a response to the challenges of AI governance in advertising, which combines elements of external (regulatory and governmental) and internal (service providers such as online platforms) governance. In this case, the regulatory model will manifest in public repositories of advertisements and regulatory control. In contrast, the self-regulatory model will manifest itself in risk management, defining the principles of corrective action and creating codes of practice.

When discussing other areas where the hybrid AI governance model can be applied, it is worth considering the labour sector. AI systems used in recruitment processes, employee performance analysis, and team management can lead to risks related to discrimination, breach of privacy, and lack of transparency in decision-making processes. For example, the Swedish system, Swedish Public Employment Services (PES), is worth mentioning. Its goal is to connect job seekers and employers effectively and contribute to long-term employment. Labour market policy evaluations are no longer carried out manually by employees, but mainly using a statistical profiling tool (Berman, 2024). The hybrid model in the labour sector means that, in regulatory terms, data protection authorities and employment agencies are authorised to supervise compliance with privacy and equality regulations in recruitment processes and employee management to prevent discrimination based on race, gender, age, or other characteristics. The self-regulatory model in this context means that companies using AI in recruitment must introduce procedures to control recruitment algorithms to ensure fairness and transparency. They should also provide accountability for decisions made by AI, including the ability to explain them to candidates.

Another example is the transport sector, including aviation. In Europe, the European Union Aviation Safety Agency is responsible for certifying and approving aviation systems (The Agency." European Union Aviation Safety Agency, 2024). For certification, EASA defines the regulations that the system developer must meet. Therefore, to enable the use of AI-based systems in aviation, EASA, in cooperation with other stakeholders, is currently developing a framework of guidelines to ensure the reliability of AI-based systems (Artificial Intelligence Roadmap 2.0, Human-centric approach to AI in aviation, 2023). As emphasised in the literature, the human factors framework for AI considers the human factors associated with its use, and the framework for mitigating the security risks of artificial intelligence introduces steps to deal with the uncertainty of artificial intelligence (EASA Concept Paper: Guidance for Level 1 &2machine learning applications, A deliverable of the EASA AI Roadmap, 2024). In this case, external (regulatory) supervision is carried out by government agencies and road safety authorities, such as the European Aviation Safety Agency (EASA) or national transport agencies. They will approve new technologies and monitor their performance in real-world conditions. In the case of self-regulatory governance, for example, manufacturers of autonomous vehicles must conduct their safety tests and audits, develop procedures for liability in the event of accidents, and apply transparency rules regarding decisions made by the AI systems in these vehicles.

A further sector worth noting is digital agriculture. Artificial intelligence can help ensure sustainable production in agriculture by streamlining farming operations and decision-making. However, improper design and configuration of intelligent systems can result in damage and unintended consequences for digital agriculture. Examples of challenges in digital agriculture include: violation of farmers'privacy, destruction of animal welfare due to robotic technology, and lack of responsibility for issues arising from the use of AI tools (Dara et al., 2022).

The hybrid supervision model is applicable in many areas where AI systems significantly impact people's lives and society. These include health, finance, transport, education, work, and agriculture. A key element in these cases is a combination of external (regulatory) and internal (self-regulatory) governance, which ensures more comprehensive risk management and compliance with applicable regulations and social values. In conclusion, it seems reasonable to assume that codes of practice are a self-regulatory model's primary form of governance. They ensure flexibility by allowing AI suppliers to determine how they will meet regulatory obligations. They can also be created by AI sector organisations, which will encourage the development of industry standards. This self-regulation can occur under EU supervision, which can determine the dissemination and development of good practices in individual sectors. Consequently, it is possible to assume that the self-regulatory governance model is based on an industry initiative, but with an EU control mechanism. This means that tech companies have the opportunity to remain autonomous in their activities, but they must act responsibly to avoid regulatory interference.

5. CORE AREAS OF GOVERNANCE IN AI

Given the subject of this study, it is essential to emphasize that issues concerning supervision in AI management encompass both internal and external areas. Regarding internal regulation, it is crucial to establish mechanisms that ensure the implementation of AI systems in compliance with rules at both the European and national levels. In the external area, on the other hand, AI system providers have a responsibility to ensure transparency, explainability, and accountability, as well as to strive to limit algorithmic bias and ensure the security of personal data. This, in turn, requires the development of information policies, e.g., indicating how users can assert their rights in the event of erroneous decisions by the AI system. This area is primarily addressed to users of the AI systems offered. In the AI Act, the EU legislator employs terminology commonly used in the scientific literature on XAI, such as transparency, opacity, and comprehensibility. In practice, this means that AI systems must be developed to ensure sufficient transparency and enable users to interpret their results correctly (Panigutti, 2023). When analysing the area related to ensuring transparency and explainability, it is essential to emphasise the obligation to provide users with clear information and rules of AI system functioning, as well as mechanisms for explaining decisions made by the AI system. Providing remedies for entities affected by AI decisions (e.g., in banking or recruitment systems) is also necessary. AI systems must be interpretable, showing how decisions are made and what factors are considered (Ausloos, 2020). With the development of autonomy and learning capabilities, modern artificial intelligence is increasingly developing algorithmic models. It generates results that are only understandable to a narrow group of specialists, remaining opaque to others, and in some cases, even impossible for society to interpret. The complexity concerns the algorithms responsible for autonomy and machine learning, as well as the contexts of their application, which are characterized by increasing diversity and complexity (Berente et al., 2021).

Transparency is therefore a key principle of ethical AI, which plays a vital role in building trust. It should be equated with openness in policies, actions, and regulations, which fosters better communication, common understanding, and cooperation. This is a foundation present in many frameworks for AI ethics. For example, according to the OECD's transparency principle (Ryan, 2020), Individuals should be informed about their data collection, participation in automated AI-based decision-making processes, and the decisions made about them resulting from the use of this data. The inability to interpret and understand the results of an AI model constitutes a violation of the principle of transparency (Jobin, 2019). Transparency means that AI systems are designed and used in a way that ensures they are traceable and explainable. This includes informing users that they are dealing with AI and providing relevant information to those who use these technologies about their capabilities and limitations. Furthermore, those affected by the AI should be aware of their rights in this context (recital 27 of the AI Act).

The term 'interpretable artificial intelligence' refers to the extent to which humans can understand the functioning and results of AI algorithms (Slack, 2019). Consequently, the transparency of AI systems should be equated with the availability of information, the rules for its provision, and how this information can practically or epistemically support the user's decision-making process (Mittelstadt,

2016). Transparency ensures that the artificial intelligence system provides information about its decision-making processes, allowing its beneficiaries and users to understand its performance and limitations by explaining, interpreting, and reproducing its decisions (Mohseni, 2021). Methods for making artificial intelligence systems transparent and comprehensible can be divided into three categories: pre-modelling (explaining data sets), internal modelling (creating models that can be interpreted), and post-modelling (building surrogate models) (Kaur, Uslu et al., 2022). Increasing transparency in AI management is intended not only to improve user confidence but also to enable the identification and elimination of potential errors or unethical activities. An important aspect is ensuring accountability and monitoring, including verifying compliance with applicable regulations, which may involve self-regulation by organisations and external oversight mechanisms established by the EU or national legislators. In conclusion, it is worth emphasising that transparency plays a fundamental role in the development of AI technology, especially in the face of the growing autonomy of systems and their applications in many fields, such as medicine, finance, and law, where errors in decision-making can have serious legal and social consequences.

The issue of explainability is closely related to the XAI mechanism. It is worth noting that in recent years, the scientific community has addressed the topic of black-box AI models (i.e. models that are too large or too complex from a technical point of view to be understood by humans), focusing on developing techniques to explain the decision-making processes of these models in a way that is understandable to humans, creating the field of eXplainable AI (XAI) (Doshi-Velez, Kim, 2017). The term 'black-box AI' refers to artificial intelligence systems whose mode of operation is difficult or impossible for humans to fully understand. This concerns advanced models that process vast amounts of data and make decisions based on complex patterns and relationships. Due to their complexity, users cannot trace the logic behind a given result, which raises challenges related to explainability, accountability, and potential algorithmic biases (Panigutti, 2023). These systems are characterised by a lack of complete transparency, which makes it difficult to analyse their operation. In technical literature, the terms 'explainability' and 'interpretability' are often used interchangeably. However, it is worth noting that their meanings may vary depending on the context (Adadi & Berrada, 2018). They refer to the extent to which it is possible to comprehend the structure of the AI model, its results, and the decision-making processes it employs. The High-Level Expert Group on Artificial Intelligence (HLEG), appointed by the European Commission, developed guidelines on the ethics of artificial intelligence in April 2019 (High Level Expert Group on Artificial Intelligence, 2019), according to which explainability means 'the ability to explain both the technical processes in the AI system and the decisions made by humans in the context of its applications'. However, it is worth noting that the question of what constitutes a 'good explanation' that is understandable to humans remains the subject of scientific debate (Miller, 2019). This is mainly due to the fact that there is no clear, generally accepted definition of a 'good explanation' in the context of AI. This is important because different fields may have different expectations regarding transparency and how AI decisions are presented. In the regulatory and ethical context, the lack of a clear definition can hinder the implementation of appropriate legal and technological standards, which is why this topic remains a key area of research and discussion.

Another area of supervision in AI management is data protection and user privacy (Okonkwo, 2020). According to recital 27 of the AI Act, privacy protection and data management involve developing and using AI systems in accordance with data protection regulations, while ensuring that information processing processes meet stringent quality and integrity standards. This issue is crucial for oversight in AI management, primarily because AI system operations require data, often including personal data. The large datasets collected in the systems can compromise the privacy of individuals whose data is disclosed. Data privacy protects individuals from being identified or linked to specific information, as automated decision-making systems can risk violating the privacy of data subjects by exposing their information to potential unauthorised disclosure or use (Mökander, Morley et al., 2021). The obligation to maintain privacy guarantees against unauthorised transfer of information, which would constitute data processing contrary to the purpose for which it was collected. From the perspective of this study, it should be emphasised that privacy issues include the challenges and values associated with data protection and the safeguarding of private and personal information (Jobin et al., 2019). Data privacy should be considered to protect individuals from being identified or linked to specific pieces of information. Automated decision-making systems can expose decision-makers to violations of the decision (Mökander & Morley, 2019; Weder et al., 2022). Privacy protects the data, including sensitive

data, provided by an individual or collected by an artificial intelligence system, safeguarding it from unjustified or illegal collection and use (Kaur, 2022). According to Article 22 of the GDPR, everyone has the right to protection against automated decision-making, including profiling. The AI Act develops and supplements these principles by introducing more detailed regulations for AI systems, particularly those with high risk. For this reason, it should be emphasised that using AI in terms of privacy requires the assumption that privacy ensures that sensitive data provided by an individual or collected by an AI system is protected from unjustified or illegal data collection and use. This is because they can be misused and have harmful consequences. Furthermore, according to Article 22 (4) GDPR, making decisions based on special categories of personal data (e.g., racial origin, political opinions, biometric data) is prohibited unless certain conditions are met. The AI Act supplements this provision by specifying additional data protection safeguards for the use of AI.

When assessing the relationship between the AI Act and the GDPR, it is crucial to recognise the complementary nature of the two acts. The AI Act expands and clarifies the principles outlined in Article 22 of the GDPR, particularly regarding the transparency of AI systems, human oversight, and the possibility of appealing decisions made by algorithms. Together, they form a legal basis that limits the use of black-box AI and ensures greater control over the impact of artificial intelligence on the lives of citizens. Other countries are also taking steps to regulate the use of black-box AI. For example, the United States has published the Blueprint for an AI Bill of Rights (Blueprint for an AI Bill of Rights), emphasizing the importance of transparency and explainability of AI systems. However, it is not a binding legal act. China, on the other hand, has introduced regulations requiring AI systems to be auditable and subject to evaluation before implementation, aiming to increase transparency.

In summary, EU legislators aim to limit the use of non-transparent AI models by designing the AI Act and the GDPR. To achieve this goal, it imposes obligations related to transparency and explainability. There is no doubt that these measures aim to ensure that AI systems operate in a way that is understandable to users and are subject to appropriate control.

6. THE IMPORTANCE OF ETHICAL PRINCIPLES FOR EFFECTIVE GOVERNANCE IN AI

As indicated, the rapid development of AI has motivated many social groups to debate its ethics. This topic is very popular among philosophers, engineers, politicians, and lawyers today. The conclusions that follow from these deliberations reach ordinary users who utilize the ever-evolving possibilities of digital tools every day (Rudnicka et al., 2020). Various organisations and institutions are increasingly addressing the issue of ethics in AI. These include national and international organisations, governmental and non-governmental organisations, as well as representatives from the commercial sector. By April 2020, these entities had issued 167 guidelines in this context. The proposed guidelines, on the one hand, are the result of the rapidly developing and ubiquitous AI. On the other hand, they carry many errors, risks, and limitations. They result from AI not creating intelligent behaviour, but only imitating it (McAfee, 2016).

The dynamic development of artificial intelligence creates enormous opportunities for many areas of life. However, rapid growth also brings ethical challenges. Effective management of emerging challenges will protect humanity, particularly in the areas of privacy and data protection. Any data obtained by AI must be processed securely to prevent unauthorized access and the unethical or illegal use of the data (Skrzecz, 2024). Therefore, standardising and regulating international ethical principles is necessary for proper and adequate supervision in AI management. This includes both the creation of a regulatory framework for AI and the implementation of audit and compliance systems, including ethical tests. Outsourcing tasks to AI can generate several abuses, the elimination of which will be a moral challenge (Taddeo and Floridi, 2016). To meet these challenges, it will be necessary to conduct an audit procedure based on ethics, i.e., capAI (Mökander & Axente, 2016). It enables the most precise compliance assessment by identifying abuses and modifying unethical behaviour. In addition, it provides in-depth knowledge of ethical considerations when designing similar systems. Importantly, the AI Act outlines the principles for conducting an audit (Mökander, Axende et al., 2021). Large-scale, ethically based audits of this kind will make it possible to avoid AI failures or improve the chances of successful system repair. In addition, they will clearly explain the causes of the failure. A correctly

performed CAI audit will enable organisations to design better, less failure-prone artificial intelligence systems. The functioning of capAI is straightforward and very effective. Thanks to the control, it is possible to eliminate the most common failure modes.

When analyzing the issue of ethics in effective AI management, it is crucial to emphasize that building public trust in AI is essential. As indicated, the Ethics Guidelines for Trustworthy Artificial Intelligence were published in April 2019. This study presents the desired perspective for developing and applying artificial intelligence in EU countries. In brief, it seeks to ensure that the development of AI in Europe respects democracy, the rule of law, and fundamental rights. Citizens should feel confident that they can safely use AI tools without risk of harm from the system (Chojnowski, 2025). The trustworthy AI has three characteristics that must be present simultaneously throughout the system's entire life. These are: legality, ethics, and reliability.

As far as legality is concerned, as the name suggests, AI should comply with generally applicable law, i.e., legislation and implementing acts. There is also no doubt about ethics; in this case, the functioning of AI should always follow ethical principles and values. The third characteristic – reliability – may raise doubts. This characteristic should be interpreted in two dimensions, both technical and social. After all, abuse can also occur when AI is used in good trust (Ethics guidelines for trustworthy artificial intelligence, 2019). So, it remains to be seen how to strengthen public confidence in AI. The answer to this question is clear – the regulatory framework for AI still needs to be strengthened. Regulations should always prioritise the safety of AI users and should be comprehensive and uniform. A milestone in this direction has already been achieved, namely with the entry into force of the AI Act.

At this point, it is also worth mentioning the Organisation for Economic Co-operation and Development (OECD) again, which laid the foundations for AI governance in 2019. According to the OECD, artificial intelligence can address challenges in education, healthcare, science, and climate action. At the same time, the organisation believes that AI poses several risks, especially to its users. These risks can be mitigated through effective governance that is based on legal regulations. In addition, the International Organization for Standardization has published ISO/IEC 42001:2023, a document designed to support organizations in using artificial intelligence systems. The document outlines principles and values to ensure that AI is developed and deployed in the best interest of society (ISO/IEC 42001:2023, 2023). The accepted standards promote the ethical and lawful use of AI through: transparency and explainability, fairness and non-discrimination, privacy and data security, responsibility and accountability, stakeholder engagement, continuous improvement, and learning. Artificial intelligence offers numerous benefits, including increased efficiency, task automation, enhanced decision-making, and support for data review. However, misused AI can cause numerous problems, including incorrect decisions, privacy breaches, bias, fraud, and cyber threats. Knowledge of the most common mistakes and the use of proven methods make it possible to limit or avoid the risks involved altogether.

The use of AI carries the risk of undesirable events. These include: a) incorrectly formulated prompts; b) data security; c) the source of the application; d) incorrect answers; and e) deepfakes and fraud.

Promptly, the questions that the user enters into the system are. The more precisely the question is asked, the greater the chance the AI will answer correctly. When using free AI systems, it is also important to remember that users pay for access with the data they enter into them, which directly leads to the risk of data security breaches. This data can be leaked and used by artificial intelligence for learning. In the context of the application's source, it is essential to emphasize that its originality should be verified each time. The justification for this practice is that there are many counterfeit platforms on the web, the use of which can pose a significant threat to their users. Fake applications are usually free and deceptively similar to the original. Therefore, the source of the application and the opinions of other users should be verified each time.

Additionally, incorrect answers are a significant risk. Applications that utilize artificial intelligence typically leverage open data sources. Therefore, the answers provided may lead to bias, prejudice, or preconceptions. As a result, the answers may be based on stereotypes that have no basis in reliable knowledge. Ultimately, AI-based applications offer numerous benefits. They support learning, work, and entertainment. However, it is essential to remember that they are often a source of misinformation or cyberattacks. The level of security in the world of AI will be significantly higher if users of artificial intelligence take the initiative to educate and train themselves. They should also be careful and verify

the displayed content, especially when the decisions have financial consequences. The role of artificial intelligence is to support and assist people, but it can never replace thinking. The careless use of artificial intelligence can lead to data leaks that pose serious risks to individuals and businesses.

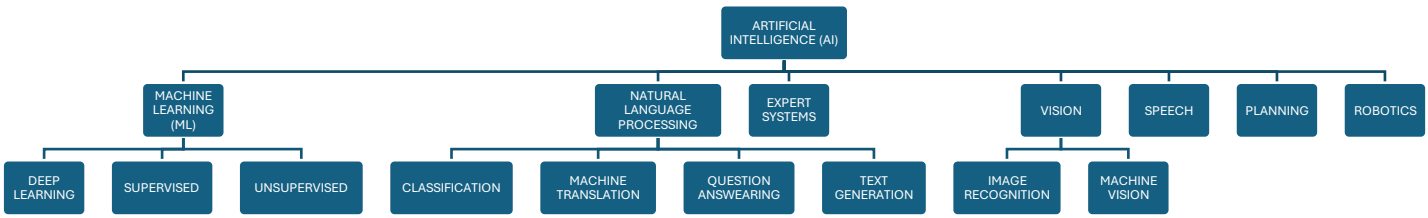
7. CHALLENGES IN COMPLIANCE AND AUDIT IN AI MANAGEMENT

The development of artificial intelligence continues to accelerate. New and more advanced tools are being developed, and existing systems on the market are continually evolving. Rapid development opens up new possibilities but also causes many risks. Therefore, today's times require constant adaptation to technology and its skilful and safe use. In addition, it is necessary to verify all legal regulations, especially those concerning personal data protection. Hence, the rapid development of AI poses a significant challenge to compliance in its broadest sense. Compliance specialists will be forced to develop new internal procedures. And to do so, they must demonstrate extensive knowledge of artificial intelligence and creativity. Verifying whether AI tools constantly comply with the law will be necessary.

On the other hand, internal practices and procedures should be continually adapted to the evolving legal environment and kept pace with the rapid development of AI. Only then will compliance officers be able to meet this challenge. At the same time, the AI audit should be kept in mind, which will enable verification of whether artificial intelligence tools function properly and whether they are unbiased and do not discriminate.

AI is the most transformative technology of our time. And it is not a single technology, but a complex set of technologies that can be combined to perform a wide range of tasks, from understanding language to helping you plan your holidays. Due to the limited scope of this study, it would be tough to discuss all technologies in an accessible and understandable way. For this reason, three basic types of AI will be analysed.

Firstly, narrow artificial intelligence only performs specific, simple tasks or sets of functions that can be found in everyday applications. It is called narrow AI because its scope of activity is limited and does not go beyond the indicated domain. Examples are facial recognition and voice assistants, which many companies and institutions already use. Secondly, there is general artificial intelligence. Its wide



range of possible actions characterises this. In principle, it can master any difficult intellectual task and perform better than a human being.

Furthermore, it can learn from experience and patterns. General AI remains a subject of ongoing research. Thirdly, there is super-intelligent AI. If it were to be developed, it would far exceed human intelligence. It could improve itself and make decisions based on data analysis. Super-intelligent AI would undoubtedly change the world. It would be able to replace humans in most processes. The following image illustrates the complexity of AI technology.

Figure 1. Artificial Intelligence (AI) Structure and Sub-Areas
Source: Raport Eurofound (2020), Game–changing technologies: Transforming production and employment in Europe, Publications Office of the European Union, Luxembourg.

European Commission documents indicate that AI as a scientific discipline has many techniques at its disposal, such as machine learning (deep and reinforcement learning), machine reasoning (planning,

reasoning, searching and optimising), robotics (control, perception, integration of other techniques) (Artificial intelligence act management and enforcement). On the other hand, OECD documents suggest that artificial intelligence can impact the environment by providing recommendations or predictions in this context. To this end, it uses machine and human input data, which enables: 1) perceiving real or virtual environments, 2) summarising such perceptions in models manually or automatically, 3) using model interpretations to formulate result options. OECD documents indicate that artificial intelligence can interfere with the environment by presenting recommendations or predictions in this context. To this end, it uses machine and human input data, which enables: 1) the perception of real or virtual environments, 2) the summarisation of such perception in models, either manually or automatically, and 3) the use of model interpretation to formulate result options.

AI is on a rapid development path and is finding applications in an increasing number of areas of life, including education, transportation, medicine, and industry. Artificial intelligence will undoubtedly play a key role in process optimisation and decision-making. It will also make it easier to diagnose diseases. These are just a few examples of the areas of life that AI will affect, and there will be many more. All this is thanks to advancements in machine learning, an area of AI that is experiencing rapid growth. This will enable even more accurate data analysis and decision-making. Artificial intelligence also brings great opportunities for the country's development. Therefore, in 2020, Poland adopted its AI development policy. The document focuses on actions and objectives in the short term (until 2023), medium term (until 2027), and long term (after 2027). They are based on six areas: society, innovative companies, science, education, international cooperation, and the public sector. This division was argued on the basis that AI permeates many areas of life; hence, the need for policy to cover all social groups as widely as possible.

With the development of artificial intelligence and its increasing use in various organisations and institutions, the need has arisen to regulate AI management standards. The ISO/IEC 42001 standard was developed in December 2024 in response to this need. It is an international standard specifying the requirements to be met when designing, implementing, maintaining, and improving an AI management system. It is intended for organisations that provide or use AI-based products and services. It is the world's first standard to regulate the management of artificial intelligence. It contains essential recommendations for the rapidly developing AI, which support organisations in managing risks and opportunities in artificial intelligence (ISO/IEC 42001:2023, 2023). It is part of a broader set of standards for implementing and improving AI. Information security management is regulated by ISO 27001, and risk management is addressed in ISO 27005 (Wygodny, 2024).

In addition, in 2024, the National Institute of Standards and Technology in the United States issued the AI RMF Generative AI Profile (NIST AI 600) under Executive Order (EO) 14110, aimed at promoting safe and trustworthy artificial intelligence to support organisations in risk management. However, on 20 January 2025, President D. Trump issued two executive orders on artificial intelligence after taking office. One of them, entitled ‘Initial Rescissions of Harmful Executive Orders and Actions’, revoked the aforementioned Executive Order (EO) 14110 (Trump Administration, 2025). It is therefore questionable whether the NIST AI 600 document will continue to be respected or whether the Trump administration will propose new solutions. These activities confirm that, despite the dynamic development of artificial intelligence, risk management is still regulated in various documents with standards that differ significantly from one another. The lack of uniform standards for AI management can lead to unforeseen risks and complications that could be eliminated by a sound management system based on uniform guidelines.

Regarding the balance between innovation and regulation in AI management, it is worth noting that the world's largest economies are vying for primacy in AI. This means that Europe must remain an active player in this field. Therefore, despite lagging behind the United States and China in technological development, the European Union has undertaken the challenging task of establishing a regulatory framework for AI. These efforts are now showing promising progress. The AI Act is the world's first comprehensive document to regulate issues related to artificial intelligence. However, it is a significant challenge to maintain innovation in mind, despite extensive regulation, and not to lose balance between the two. There is no shortage of concerns in this context. Organisations indicate that EU regulations could hamper artificial intelligence development, exceptionally if restrictions on technical stability and liability for damage caused by artificial intelligence providers are maintained.

8. CONCLUSIONS

Modern artificial intelligence systems are playing an increasingly important role in various sectors, which necessitates the implementation of appropriate supervisory mechanisms and adherence to ethical principles. Key areas of AI governance include ensuring transparency and accountability, minimising algorithmic bias, protecting privacy and personal data, and auditing and certifying AI systems. From a regulatory perspective, global initiatives such as the European Union's AI Act, the Blueprint for an AI Bill of Rights developed in the USA, and the restrictive AI regulations in China are significant. Technical solutions that support an ethical approach to artificial intelligence, such as Explainable AI (XAI), and mechanisms that eliminate so-called 'black boxes' in AI models, also play an essential role. Scientific discussions continue to focus on the issues of explainability and interpretability of AI systems, which are crucial for building public trust and the successful implementation of artificial intelligence in critical areas such as healthcare, the financial sector, and the justice system. The regulations governing this area are designed to ensure that AI functions adhere to the principles of ethics, human rights protection, and algorithmic justice.

Creating effective legal regulations regarding ethics and governance in AI management is crucial. For this reason, it is necessary to distinguish among three governance models in this publication: regulatory, self-regulatory, and hybrid. Legal regulations created at the EU and national levels (regulatory supervision model) are of fundamental importance in ensuring compliance with the law of their creation and use, as well as in protecting fundamental rights. The use of AI systems in various sectors of everyday life has a direct impact on human rights. Artificial intelligence can affect the fundamental rights and freedoms of individuals, including a) the right to privacy and the protection of personal data (AI is often based on the processing of vast amounts of data, which carries the risk of misuse, profiling of users or violation of the GDPR); b) the right to equal treatment (algorithms learn from historical data, which may reflect existing prejudices in society; AI systems not covered by any regulations may perpetuate or even exacerbate inequalities, e.g. in recruitment processes or credit decisions); c) the right to a fair trial (AI is increasingly used in justice systems, e.g. for simple organisational and technical activities). Given the above, it is reasonable to assume that the legal framework governing the operation of AI should ensure that decisions made using AI are transparent and subject to human oversight and control. In addition, creating clear regulations ensures that AI systems operate according to democratic values and fundamental rights.

The second area of importance from the perspective of regulating the ethics and governance of AI management is security. Numerous risks associated with the use of AI-based systems are being identified today. With the systematic implementation of AI in key sectors such as healthcare, justice, and finance, the need for regulation regarding responsibility for decisions made by algorithms is growing. In this context, the main risks are, for example, wrong medical decisions when using AI in the diagnosis and treatment of patients. A lack of regulation can lead to a situation where it is unclear who is responsible for an incorrect diagnosis generated by an algorithm.

Another example is autonomous emergency systems, as autonomous vehicles or robots can cause accidents for which liability is not clearly assigned (i.e., manufacturer, operator, or user?). Additionally, cybersecurity issues must be considered, as AI systems are increasingly vulnerable to hacking. This, in turn, can lead to data leaks or the manipulation of decisions made by algorithms.

An analysis of current and proposed legal regulations on ethics and governance in AI management has led to the conclusion that transparency in AI is essential. Therefore, citizens, users, and institutions must be able to trust that decisions made by AI are comprehensible, open to challenge, and subject to human oversight and control. This means that the system's users should be aware of the basis on which the algorithm made a decision (e.g., when assessing creditworthiness), and regulations should ensure the right to appeal against AI decisions, especially in cases involving important issues. Importantly, AI cannot act completely autonomously in situations affecting citizens' rights. The introduction of legal regulations at the EU level, obliging member states to create internal rules and AI system providers to establish their own internal procedures, currently favours ensuring an adequate level of AI transparency. The development of their own ethical codes for AI by providers will minimise legal risk, build trust among customers and employees, and increase the credibility of organisations through transparency. These measures are complemented by so-called soft law acts, which support the regulatory and self-regulatory governance model. The coherence of regulations at the EU level is noticeable, as exemplified

by the principles formulated by the EU legislator in the GDPR, particularly the principle of explainability, which affords everyone the right to information about how algorithms that affect their lives operate. The AI Act also imposes a transparency obligation on high-risk systems.

The analysis of the current legal regulations shows that the European Union is currently the leader in AI regulation. The creation of a uniform AI regulation system at the EU level is crucial. First of all, it prevents legal fragmentation. In the absence of harmonised regulations, different regulations in individual countries could hinder the functioning of entrepreneurs and institutions in the single market. Secondly, it facilitates business activities, as it is possible to comply with a single EU standard instead of having to adapt to different national regulations. Thirdly, it is about protecting democratic values. The activity of the EU legislator ensures that AI regulations reflect the European approach to protecting human rights and promoting technological ethics.

In summary, well-designed regulations should not limit the development of AI but provide a legal framework for innovation. In an era of leveling geographical barriers and shifting a significant portion of activity to the online sphere, it is essential to maintain interoperability standards. This facilitates cooperation between different AI systems in Europe and worldwide. Regulations on artificial intelligence are necessary to ensure security, transparency, and the protection of citizens' rights. At the same time, well-designed laws can promote innovation and competitiveness. For this reason, the process of harmonizing regulations at the EU level, adapting them at the national level, and implementing ethical principles within organizations should be assessed as unequivocally positive.

REFERENCES

Adadi A., Berrada M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE access PP*. 6. 5213852160. https://doi.org/10.1109/ACCESS.2018.2870052.

AI Ethics Guidelines Global Inventory. Retrieved from: https://inventory.algorithmwatch.org (05.03.2025).

Akerkar, R. (2019). *Artificial Intelligence for Business*. Cham: Springer.

Ausloos J., Veale M., (2020). Researching with Data Rights. *Amsterdam Law School Research Paper*. No. 2020–30, 136–157.

Badulescu D., Simut R., Bodog S. A., Badulescu A., Simut C., Zapodeanu D. (2025). Shaping AI-related competencies for the labor market and business. A PLS-SEM approach, *International Journal of Computers, Communications & Control*. 20(1). https://doi.org/10.15837/ijccc.2025.1.6894.

Bietti E. (2020, January). From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. *Conference: FAT* '20: Conference on Fairness, Accountability, and Transparency.* 210–219. https://doi.org/10.1145/3351095.3372860

Blueprint for an AI Bill of Rights. Retrieved from: https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/what-is-the-blueprint-for-an-ai-bill-of-rights/ (05.03.2025).

Berente N., Gu B., Recker J., Santhanam R. (2021). Managing artificial intelligence. *MIS Quarterly*. *45*(3). 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Berman A., de Fine Licht K., Carlsson V. (2024). Trustworthy AI in the public sector: An empirical analysis of a Swedish labor market decision-support system. *Technology in Society*. *76*. 102471. https://doi.org/10.1016/j.techsoc.2024.102471.

Cebrian M., Gómez E., Fernández-Llorca D. (2025, January 10). Supervision policies can shape long-term risk management in general-purpose AI models. https://doi.org/10.48550/arXiv.2501.06137.

Chojnowski M. (2025). Jak używać sztucznej inteligencji zgodnie z wytycznymi Komisji Europejskiej w zakresie etyki? Retieved from: https://www.gov.pl/web/ai/godna-zaufania-ai-czyli-jak-uzywac-sztucznej-inteligencji-zgodnie-z-wytycznymi-komisji-europejskiej-w-zakresie-etyki (05.03.2025).

Dara R., Hazrati Fard S. M., Kaur J. (2022, July). Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Frontiers in Artificial Intelligence*, 5. 5:884192. https://doi.org/10.3389/frai.2022.884192.

Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119, 4.5.2016. 89–131.

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), OJ L 201, 31.7.2002. 37–47.

Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC. PE/7/2024/REV/1. OJ L, 2024/2853. 18.11.2024.

Doshi-Velez F., Kim B. (2017, March 2). Towards a rigorous science of interpretable machine learning. Retrieved from: http://arxiv.org/abs/1702.08608 (05.03.2025).

European Declaration on Digital Rights and Principles for the Digital Decade. COM/2022/28 final

European Union Aviation Safety Agency (2024). "The Agency." European Union Aviation Safety Agency, Ed., European Union Aviation Safety Agency. Retrieved from: https: //www.easa.europa.eu/en/the-agency/the-agency (20.03.2025).

European Union Aviation Safety Agency (EASA) (2025, March). "Artifical Intelligence Roadmap 2.0, Humancentric approach to ai in aviation," European Union Aviation Safety Agency (EASA), Postfach 10 12 53, 50452 Cologne, Germany, Tech. Rep.., version 2.0, 2023-05.

European Union Aviation Safety Agency (EASA), "EASA Concept Paper: Guidance for Level 1 & 2 machine learning applications, A deliverable of the EASA AI Roadmap," European Union Aviation Safety Agency (EASA), Postfach 10 12 53, 50452 Cologne, Germany, Tech. Rep., version Issue 02, 2024-03-06. Retrieved from: https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial intelligence-concept-paper-issue-2 (05.03.2025).

Farayola O.A., Abdul A. A., Irabor B. O., Okeleke E. Ch. (2023). Innovative business models driven by AI technologies: a review. *Computer Science & IT Research Journal*, *4*(2), 85-110. https://doi.org/10.51594/csitrj.v4i2.608.

Gunning D. (2017). *Explainable artificial intelligence* (xai). Defense Advanced Research Projects Agency (DARPA).

High Level Expert Group on Artificial Intelligence (2019). Ethics Guidelines for Trustworthy AI.

Holzinger A., Saranti A., Molnar Ch., Biecek P., Samek W. (2022). *Explainable AI methods: a brief overview. In: International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. Springer, 13–38. https://doi.org/10.1007/978-3-031-04083-2_2.

ISO/IEC 42001:2023. (2023). Retrieved from: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:42001:ed-1:v1:en (05.03.2025).

Jobin A., I. Marcello, V. Effy, The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9). 389–399. https://doi.org/10.1038/s42256-019-0088-2 (05.03.2025).

Kaur D., Uslu S., Rittichier K.J., Durresi A. (2022). Trustworthy Artificial Intelligence: a review, *ACM Computing Surveys*. 55(2). 1–38. https://doi.org/10.1145/3491209.

Kroll J., Huey J., Barocas S., Felten E., Reidenberg J., Robinson D., Yu H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*. *165*. 633–705.

Korzyński P., Costa e Silva S., Górska A.M., Mazurek (2024, September 12). Trust in AI and Top Management Support in Generative-AI Adoption, *Journal of Computer Information Systems*, https://doi.org/10.1080/08874417.2024.2401986.

Laine J., Minkkinen M., Mäntymäki M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders, *Information & Management.* 61. https://doi.org/10.1016/j.im.2024.103969.

La Brie R., Steinke G. (2019). Toward a framework for ethical audits of AI algorithms, In: *Twenty-fifth Americas Conference on Information Systems*. 33–44

Maciążka A. (2025, March). Sezon na AI: Rozporządzenie o sztucznej inteligencji. Retrieved from: https://www.gkrlegal.pl/ai-summer-ustawa-o-sztucznej-inteligencji/ (05.03.2025).

McAfee, A., Brynjolfsson, E. (2016). *The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant*. Technologies, New York and London. https://doi.org/10.1080/14697688.2014.946440.

Miller T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*. *267*. 1–38. https://doi.org/10.1016/j.artint.2018.07.007.

Min, H. (2010). Artificial intelligence in supply chain management: theory and ap- plications. *International Journal of Logistics: Research and Applications*, *13*(1), 1339.

Mittelstadt B., Allo P., Taddeo M., Wachter S., Floridi L. (2016). The ethics of algorithms: mapping the debate. *Big Data & Society In press*. *2*. 1–21. https://doi.org/0.1177/2053951716679679.

Mökander J., Axente M. (2021, October). Ethics-based auditing of automated decision-making systems: intervention points and policy implications, with AI & Society. https://doi.org/10.48550/arXiv.2110.10980.

Mökander J., Axente M., Casolari F., Floridi L. (2021). Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation. *Minds and Machines*. *32*. 1-27. https://doi.org/10.1007/s11023-021-09577-4.

Mökander J., Morley J., Taddeo M., Floridi L. (2021, October). Ethics-based auditing of automated decision-making systems: nature, scope, and limitations. *Science and Engineering Ethics*. *27*(4). 44. https://doi.org/10.1007/s11948-021-00319-4.

Mohseni S., Zarei N., Ragan E.D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems, ACM. *The ACM Transactions on Interactive Intelligent Systems*. *11*(3–4). 1–45. https://doi.org/10.1145/3387166.

Naron-Grochalska E. (2024). AI Act: nowe ramy prawne dla sztucznej inteligencji w Europie. Retrieved from: https://getsix.pl/getsix-blog-pl/podatki-i-prawo-alerts-polska/podatki-i-prawo/ai-act-nowe-ramy-prawne-dla-sztucznej-inteligencji-w-europie/ (05.03.2025).

Okonkwo, C. W., Ade-Ibijola, A. (2020). Python-bot: A chatbot for teaching python programming. *Engineering Letters*, *29*(1), 25-35.

Overview of all AI Act National Implementation Plans (2024, November 8). Retrieved from: https://artificialintelligenceact.eu/national-implementation-plans/ (28.03.2025).

Panigutti C., Hamon R., Hupont I., Llorca D. F., Yela D. F., Junklewitz H., Scalzo S., Mazzini G., Sanchez I., Garrido J.S., Gomez E. (2023). The role of explainable AI in the context of the AI Act. *In 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT'23), June 12–15, 2023, Chicago, IL, USA. ACM, New York,* 1139-1150. https://doi.org/10.1145/3593013.3594069.

Pournader, M., Ghaderi, H., Hassanzadegan, A., & Fahimnia, B. (2021). Artificial intel- ligence applications in supply chain management. *International Journal of Production Economics*, 241, Article 108250.

Projekt polskiej ustawy o systemach sztucznej inteligencji z 17.10.2024 r., nr UC71 (eng. Draft Polish law on artificial intelligence systems, 2024, October 17, No. UC71). Retrieved from: https://legislacja.rcl.gov.pl/lista?typeId=2&_typeId=1&title=sztucznej+inteligencji&createDateFrom=&createDateTo=&applicantId=&number=&_isUEAct=on&_isTKAct=on&_isActEstablishingNumber=on&_isSeparateMode=on&_isDU=on&_isNumerSejm=on#list (25.03.2025).

Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial, Publicado en: «BOE» núm. 210, de 2 de septiembre de 2023, páginas 122289 a 122316. Retrieved from: https://www.boe.es/eli/es/rd/2023/08/22/729 (30.03.2025).

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277, 27.10.2022, 1–102.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119, 4.5.2016, 1–88. (GDPR)

Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (Text with EEA relevance), OJ L 295, 21.11.2018, 39–98.

Rudnicka A., Kaczorowska-Spychalska D., Kulik M., Reichel, J. (2020). Digital ethics – polscy konsumenci wobec wyzwań etycznych związanych z rozwojem technologii. I Ogólnopolski Raport. Łódź. https://doi.org/10.18778/8220-413-1.01.

Ryan M. (2020). Agricultural big data analytics and the ethics of power. *Journal of Agricultural and Environmental Ethics*. 33. 49–69. https://doi.org/10.1007/s10806-019-09812-0.

Skrzecz K. (2024). Etyka w sztucznej inteligencji: jak zapewnić, że AI będzie służyć dobru ludzkości? Retrieved from: https://przemyslprzyszlosci.gov.pl/etyka-w-sztucznej-inteligencji-jak-zapewnic-ze-ai-bedzie-sluzyc-dobru-ludzkosci (05.03.2025).

Slack D., Friedler S. A., Scheidegger C., Roy C. D. (2019, February). Assessing the local interpretability of machine learning models. https://doi.org/10.48550/arXiv.1902.03501.

Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework. Retrieved from: https://www.edpb.europa.eu/our-work-tools/our-documents/statements/statement-32024-data-protection-authorities-role-artificial_en (30.03.2025).

Taddeo M., Floridi L., What is data ethics? *Philosophical Transactions of the Royal Society A*. 1-5. 20160360. https://doi.org/10.1098/rsta.2016.0360.

Taebi B., van den Hoven J., Bird S.J. (2019). The importance of ethics in modern universities of technology. *Science and Engineering Ethics, 25*. 1625–1632.

Trump Administration Issues Executive Order on Artificial Intelligence (2025). Retrieved from: https://www.sullcrom.com/SullivanCromwell/_Assets/PDFs/Memos/Trump-Administration-Issues-Executive-Order-Artificial-Intelligence.pdf (12.03.2025).

United States: AI tug-of-war - Trump pulls back Biden's AI plans. https://insightplus.bakermckenzie.com/bm/data-technology/united-states-ai-tug-of-war-trump-pulls-back-bidens-ai-plans (05.03.2025).

Urbanowicz A. (2025). Akt o usługach cyfrowych a ogólne rozporządzenie o ochronie danych: kluczowe wyzwania dla administratorów jako dostawców usług pośrednich z branży marketingu. In: *Ochrona danych osobowych w marketingu i sprzedaży*, M. Gumularz, P. Kozik (eds.). Warsaw. 385.

Werder K., Ramesh B., Zhang R. (2022). Establishing data provenance for responsible Artificial Intelligence systems, *ACM Transactions on Management Information Systems*, *13*(2). 1–23, https://doi.org/10.1145/3503488.

Wygodny A. (2024). Sztuczna inteligencja w audycie i zarządzaniu ryzykiem https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence (12.03.2025).

Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji (eng. Ethics guidelines for trustworthy artificial intelligence). (2019). Bruksela. Retrieved from: https://www.gov.pl/web/ai/etyka (05.03.2025).

Zarządzanie aktem w sprawie sztucznej inteligencji i jego egzekwowanie (eng. Artificial intelligence act management and enforcement). Retrieved from: https://digital-strategy.ec.europa.eu/pl/policies/ai-act-governance-andenforcement?utm_source=chatgpt.com#:~:text=wykaz%20wszystkich% 20zidentyfikowanych%20organ%C3%B3w%20(.XLS). (12.03.2025).