PANEVĖŽIO KOLEGIJA ISSN 2029-1280, eISSN 2669-0071. Taikomieji tyrimai studijose ir praktikoje – Applied Research in Studies and Practice, 2025, 21.

ARTIFICIAL INTELLIGENCE ACT AND CUSTOMER SAFETY IN THE BANKING SECTOR – LEGAL AND ORGANISATIONAL CHALLENGES

Iwona LASEK-SUROWIEC, Marcin OSKIERKO, Sławomir ŻURAWSKI

State Academy of Applied Sciences in Chełm, Poland

Abstract. The aim of the article is to comprehensively analyze the impact of the Artificial Intelligence Act (AI Act) on customer security in the banking sector and to identify the most important legal, technological and organizational challenges related to the implementation of artificial intelligence systems. The study used the method of systematic literature review, comparative analysis of regulatory documents and review of supervisory reports. The results indicate that AI systems used in banking – especially in the area of credit risk assessment and anti-fraud – generate new types of risk, including algorithmic risk, data bias, opacity of decision-making processes, and susceptibility of models to manipulation. The AI Act introduces obligations on data quality, auditability, human oversight and model monitoring to strengthen consumer protection. The identified challenges include a lack of specialist competence, immaturity of the data infrastructure, high implementation costs and the need to reorganise model management processes. The conclusions indicate that the effectiveness of the AI Act depends on the ability of financial institutions to adapt technologically and organizationally, as well as on building a culture of responsible and transparent use of AI in banking.

Keywords: AI Act, artificial intelligence, customer security, algorithmic risk, financial regulation

INTRODUCTION

The dynamic development of artificial intelligence and its intensive implementation in the financial sector make the security of bank customers one of the key challenges of the modern digital economy. European Union regulations, in particular the Artificial Intelligence Act (AI Act), introduce a new quality in terms of legal and organizational requirements for the design, use and supervision of high-risk AI systems, which include solutions used in banking. The importance of the problem results from the growing scale of threats – from automated cyberattacks, through abuses in AI decision-making processes, to the risk of discrimination against customers in algorithmic models of creditworthiness assessment. These issues are increasingly being analysed in the literature, including both works on technology regulation (Bryson, Floridi) and research on financial security and data protection in the banking sector (Arner, Barberis, Zetzsche).

Previous scientific studies, taking into account, m.in the development of regtech, fintech and challenges related to operational risk management, emphasize the need to combine legal, technological and organizational perspectives. At the same time, they point out that current surveillance models often do not keep up with the pace of development of artificial intelligence solutions. This analysis allows for the formulation of key research questions: how does the AI Act affect the security of financial institution customers? Do the obligations imposed on banks in the areas of transparency, data management and supervision of algorithms actually increase the level of protection? What are the legal and organisational gaps that may limit the effectiveness of the implementation of the new regulations?

Therefore, the research hypothesis of this study assumes that the AI Act will significantly increase the security standards of banking sector customers, but the effectiveness of these solutions will depend on the ability of banks to adapt organizationally, invest in technological risk management and properly implement supervisory processes over AI systems. This hypothesis is supported by numerous international studies that prove that technology regulations have a real impact on the stability of the financial system only if they are accompanied by adequate resources, competencies and a safety culture in the organization.

At the same time, the literature review, including publications indexed in international databases such as Scopus or Web of Science, indicates that the issue of AI regulation in banking is at the stage of intensive scientific development. These studies analyze both legal and ethical aspects (European Banking Authority, OECD), as well as technical issues related to algorithmic security, data quality or resilience of systems to cyber threats. This shows the high level of complexity of the problem under study and the multidimensional nature of the challenges facing the financial sector.

The originality of this article lies in its holistic view of the relationship between the AI Act and the security of bank customers, combining a scientific perspective with practical implications for financial institutions. This analysis contributes to the development of research on digital security, complementing the existing work with the context of new European regulations and their impact on the functioning of the banking sector. This provides the reader with a solid foundation for understanding the scope of the study, a review of the scientific evidence to date, and the direction in which the development of customer protection is heading in an era of increasing automation of financial processes.

THE RESEARCH METHOD

The study was designed as a qualitative-analytical comparative analysis, based on recognized and cited international literature indexed in the Scopus and Web of Science databases, including works on technology regulation (m.in. Floridi, Bryson), financial security (Arner, Barberis, Zetzsche) and institutional guidelines of the EBA, OECD and

the European Commission on the AI Act. PRISMA, which selected publications from 2018-2025 on artificial intelligence in the banking sector, algorithmic risk assessment and financial security regulation.

The study used content analysis and institutional and legal analysis. The content analysis included thematic coding of identified issues, such as: transparency of AI systems, data management, risk of algorithmic discrimination, cybersecurity, regulatory oversight and legal liability. The institutional analysis was carried out by comparing the current obligations of banks under sectoral regulations (m.in. DORA, PSD2, GDPR) with the new requirements of the AI Act.

The study did not include participants or quantitative data, only existing sources were analyzed, in accordance with the desk research method. The independent variables were: the scope of the AI Act regulation, the characteristics of AI systems in banking and the types of technologies used. The dependent variables are the level of customer security, the level of operational risk and the resistance of algorithms to breaches. No observations were rejected because the analysis concerned a full body of literature meeting the methodological criteria. The selection of the literature sample was determined on the basis of citation rates (IF, JCR, SNIP) to ensure high quality of sources.

The analytical methods used, including triangulation of sources, comparative analysis and thematic categorization, guarantee resistance to interpretation errors, enabling unambiguous and reliable conclusions to be drawn regarding the relationship between the AI Act and the safety of bank customers.

DISCUSSION/RESULTS

The analysis of the collected research material allows us to identify several key conclusions regarding the impact of the AI Act on the security of bank customers and the functioning of financial institutions. First, the new regulations unambiguously classify most AI systems used in banking – including creditworthiness assessment systems, AML transaction monitoring, behavioural scoring and fraud detection systems – as high-risk systems. This means that there is an obligation to introduce stringent requirements for transparency, data quality, accountability and auditability of models. The results of the analysis indicate that banks' current practices do not always meet these standards, especially in the area of model validation and documentation of machine learning processes.

The second important result is that the AI Act significantly strengthens customer protection through the requirement to use human oversight mechanisms and the obligation to inform about the use of AI systems in consumer decision-making processes. At the same time, the literature emphasizes that the implementation of such solutions will require significant organizational changes, m.in. expansion of compliance teams, creation of new technology risk management procedures or standardization of audit processes.

The results of the regulatory benchmarking also indicate that the AI Act complements the legal acts in force in the banking sector – such as GDPR, DORA or PSD2 – creating a more coherent legal environment regarding technological security. At the same time, however, legal gaps are emerging, particularly in the area of liability for damage caused by autonomous decision-making systems and in the standardisation of algorithmic audit methods. The scientific discussion indicates that the full effectiveness of the AI Act will depend on the harmonisation of implementation practices at the European level and on the supervisory capabilities of financial authorities.

The analysis also shows that despite the increase in regulatory requirements, the implementation of the AI Act alone does not guarantee an automatic increase in security. Investments in cybersecurity, model monitoring tools, and management training are needed. Banks that fail to adapt their procedures will remain vulnerable to incidents related to algorithmic errors or data breaches, as evidenced by numerous cases described in the literature.

To sum up, the results obtained indicate that the AI Act can significantly increase the level of security of banking sector customers, but the condition for the effectiveness of the new regulations is the appropriate organizational preparation of financial institutions and the creation of strong supervisory competencies. The discussion also confirms that further research should focus on the empirical assessment of the implementation of regulations at the operational level and on the analysis of the effectiveness of AI models in real-world threats.

AI Act and security regulations in the banking sector

The Artificial Intelligence Act (AI Act) is the first in the European Union and one of the first in the world comprehensive regulations relating to the functioning of artificial intelligence systems. It was developed in response to the dynamic development of machine learning technologies and the growing number of AI applications in areas requiring a high level of public trust, such as the banking and financial sectors. Banking, as one of the most regulated branches of services, has naturally become a field of intense discussion regarding the legal and organizational consequences of the implementation of the AI Act.

In recent years, the use of artificial intelligence in banking has accelerated intensively. Learning systems have become a key element of credit risk assessment, anti-money laundering (AML/CFT), financial fraud identification, transaction monitoring, service personalization, and customer service automation. The ability to process massive data sets and detect irregularities at speeds impossible for humans to do has significantly improved banking processes, reducing operational costs and increasing efficiency. At the same time, the growing use of AI has revealed new types of threats that have not been covered by consistent regulation until now. These include, m.in example, the risk of algorithmic errors, the susceptibility of models to manipulation, data bias, the opacity of decision-making processes, and the lack of accountability for the effects of autonomous systems (Barocas, Selbst, 2016).

The AI Act addresses these challenges by classifying AI systems according to their level of risk. The most stringent regulatory regime includes systems classified as high-risk, including almost all AI tools used by banks. The

European Commission has explicitly indicated that algorithmic credit scoring, anti-money laundering tools and consumer solvency assessment systems are among the technologies that have the potential to have a direct impact on citizens' fundamental rights and must therefore be subject to the strictest supervisory requirements (European Commission, 2023). Such a qualification results from the fact that decisions made by algorithms in the banking sector may determine access to financial services, affect the economic situation of customers, and in extreme cases lead to their financial exclusion.

The regulation introduces a number of obligations regarding the design, training and monitoring of artificial intelligence systems. One of the most important is the need to ensure high-quality data used to train models. The data must be complete, representative, up-to-date, and free from biases and biases that could influence the decision outcome. This is especially important in banking, where historical customer data often reflects social, regional or economic biases. Research indicates that many scoring models can favor some customer groups and discriminate against others, especially when they are trained on historical data without proper quality control (Kleinberg et al., 2018).

Another key requirement of the AI Act is to ensure that AI systems are fully auditable. Financial institutions must maintain documentation that allows the model's decision-making process to be reconstructed and the quality of the data, parameters and assumptions made during machine learning to be verified. This is a response to the problem of so-called "black boxes" — deep learning models that generate results that cannot be directly interpreted by humans (Samek, Müller, 2019). In the context of banking, where every lending decision must be legally justified, the lack of transparency of models has been a significant controversy for many years.

The AI Act also requires the provision of human oversight of key decision-making systems. In practice, this means that decisions generated by artificial intelligence — especially those of a negative nature or with high uncertainty — must be verified by an employee of the institution. This obligation introduces an element of human responsibility, minimizing the risk of automatic errors. The European Banking Authority (EBA, 2023) indicates that the lack of human oversight of scoring and AML systems was one of the most frequently identified problems during technology audits conducted in European financial institutions.

The document also strengthens existing sectoral regulations, such as GDPR, DORA or PSD2, creating an integrated consumer protection system with them. While the GDPR focuses on the protection of personal data, the AI Act covers the full lifecycle of a system, including how data is acquired, how models are trained, tested, implemented, monitored, and responded to errors. DORA (Digital Operational Resilience Act), on the other hand, focuses on the operational resilience of financial institutions, and the new AI regulation complements it, introducing an element of responsibility for the quality and security of decision-making processes.

However, there are numerous voices in the literature that despite its extensive obligations, the AI Act does not eliminate all the risks associated with the use of artificial intelligence in banking.) For example, liability for damage resulting from the operation of automated systems remains a challenge (Zetzsche, Buckley, 2023).

Machine learning models can make decisions that aren't obviously wrong, but lead to negative financial consequences for consumers. In such cases, it can be complicated to determine the liability between the bank, the software developer and the data provider.

Another problem is the readiness of financial institutions to implement the new requirements. Most banks do not have sufficient technological infrastructure to monitor models in real time, and they also lack staff specialized in algorithmic auditing (McKinsey & Company, 2023). The requirements of the AI Act may therefore lead to increased operating costs, the need to restructure business processes and the reorganization of structures responsible for technology risk management.

Despite the difficulties, the AI Act is a breakthrough in the approach to customer safety in the banking sector. For the first time, the rules of operation of algorithmic decision-making systems have been regulated in such detail, introducing quality, supervision, accountability and transparency requirements. It can be assumed that the AI Act will become the foundation of a new era in financial supervision, in which not only legal compliance, but also the ethical and technical stability of AI models will play a key role.

Customer security and algorithmic risk in the banking sector

Artificial intelligence in the banking sector has become one of the key tools supporting analytical, decision-making and operational processes. Automating these processes has brought numerous benefits, including increased operational efficiency, reduced costs, and improved customer service. However, with the increasing use of algorithms, new forms of risk have emerged that have not been sufficiently addressed in traditional risk management frameworks so far. These risks include decision-making errors, algorithmic bias, susceptibility to manipulation, opacity of models, and risks from data quality. As a result, customer safety has become one of the main areas requiring regulatory changes, which was one of the key arguments for the implementation of the AI Act.

One of the basic risks associated with the use of AI in banking is the so-called risk of algorithmic errors. Machine learning models—especially those based on deep neural networks—can generate misclassifications, especially in situations where the input differs from the data used during training. The work of Goodfellow, Shlens and Szegedy (2015) has shown the vulnerability of neural networks to so-called adversarial examples, i.e. intentionally prepared input data that can drastically change the model's decision with minimal changes in the data structure. In the context of banking, this means that the credit scoring system or fraud detection tools can be manipulated, which can lead to serious financial consequences.

The second key area of risk is algorithmic bias. Algorithms, especially those trained on historical data, can replicate existing social inequalities and discriminate against certain groups of customers (Barocas, Selbst, 2016). For

example, credit data can favor people from certain regions, incomes, or demographics, perpetuating patterns of financial exclusion. This phenomenon is particularly dangerous in the banking sector, because algorithmic decisions concern issues fundamental to social life, such as access to credit, renting an apartment, buying real estate or the possibility of running a business.

Algorithmic risk also concerns the opacity of decision-making processes. Deep learning-based models are characterized by a black box structure, in which even the creators of the system are unable to unambiguously explain the reasons behind a particular model decision. Samek and Müller (2019) emphasize that the lack of explainability (XAI – Explainable Artificial Intelligence) is one of the key challenges from the point of view of supervision, regulatory compliance and consumer protection. In the event of a refusal to grant a loan, the client has the right to know the justification for the decision, but in the case of an algorithm acting as a "black box", such justification may be difficult to obtain.

The quality of the data used to train models is also an extremely important issue. Machine learning-based models are extremely vulnerable to data errors – incomplete data, lack of updates, human errors entered during input development, and inadequate representativeness can result in serious irregularities. A report by the Bank for International Settlements (BIS, 2023) highlights that in the financial sector, data is particularly vulnerable to systematic errors, which can lead to unfair or incorrect algorithmic decisions. One incorrect parameter in the training data can lead to errors in the classification of thousands of customers, which can result not only in financial losses, but also in a loss of public trust.

Another important issue is the operational risk resulting from process automation. The BIS report draws attention to the risk category referred to as model risk, i.e. the risk resulting from the improper operation of the model or its application in conditions other than those for which it was designed. Scoring models or AML systems can function properly in stable conditions and fail in periods of greater economic volatility. The COVID-19 pandemic and the energy-inflation crisis have shown that many predictive models in the financial sector do not cope with crisis situations when customer behavior patterns change.

The literature also points to the risks associated with cybercriminals taking control of the models. They exploit algorithm vulnerabilities, m.in through so-called data poisoning, deliberately infecting the data used to train the model. AI security research has shown that even a small amount of artificially manipulated data can cause a system to malfunction (Goodfellow et al., 2015). In the banking sector, the consequence of such actions may be the weakening of fraud detection systems, the misclassification of transaction risk or the circumvention of biometric security.

The AI Act addresses these challenges by introducing requirements for explainability, data quality, human oversight, and auditability. This is particularly important because customer protection in banking is one of the foundations of the stability of the financial system. Regulation requires that decision-making systems are not only effective, but also fair, secure and verifiable. Thanks to this, customers are to gain greater protection against technological errors, and banks are to gain tools that allow for effective risk monitoring.

The use of artificial intelligence in the banking sector brings significant benefits, but at the same time generates new, complex forms of risk that can directly affect customer safety. Algorithmic risk, data bias, vulnerability, opacity of models and operational errors pose a real threat to the stability of the financial sector. For the first time, the AI Act comprehensively addresses these threats, creating a legal framework to protect customers from the negative effects of uncontrolled AI technologies.

Organizational and technological challenges of implementing AI Act in the banking sector

The implementation of the AI Act in the banking sector is associated with the need to introduce deep organizational and technological changes, which in practice may determine the effectiveness of regulations. While the AI Act creates a coherent legal framework for the responsible use of AI, the real challenge for banks remains to adapt their structures, processes and resources to the new requirements. Scientific literature and industry reports emphasize that the degree of organizational maturity of financial institutions in the area of model, data and technological risk management still remains diverse (Zetzsche & Buckley, 2023). As a consequence, the implementation of the AI Act will require both procedural changes and far-reaching investments in technology.

One of the main challenges is the need to create a comprehensive framework for managing the lifecycle of artificial intelligence models, referred to as Model Risk Management (MRM), in banks. Most financial institutions still do not have dedicated models oversight strategies, and existing policies often do not include advanced learning algorithms (McKinsey 2023).

The AI Act, on the other hand, requires that any model classified as high-risk be subject to formal procedures including: documentation, validation testing, risk analysis, stability monitoring and updates. This means the need for new organizational structures, such as AI governance teams or specialized units responsible for testing and auditing algorithms.

At the same time, banks must ensure that they have the right technological infrastructure in place to enable the legal use of AI systems. Deloitte (2024) highlights that many financial institutions still use distributed data environments that do not provide an adequate level of control over the quality of information going into models. The AI Act requires training and operational data to be high-quality, complete, and free from bias – which requires the creation of centralized repositories, data profiling systems, and anomaly detection tools. Banks must therefore invest in modernizing their data architecture, creating internal "data labs" and implementing Data Governance standards in accordance with regulatory requirements.

Another challenge is the introduction of automated mechanisms for monitoring models. Machine learning models tend to drift, i.e. degrade the quality of predictions as economic conditions, market trends, or customer behavior

change. The BIS report indicates that the lack of automatic model drift detection systems was one of the most common problems detected during technology audits in banks. The implementation of the AI Act means the need to introduce tools that monitor the operation of models in real time and automatically report deviations. This is especially important in AML/CFT areas, where models need to detect new patterns of abuse and adapt to changing criminal behaviors.

Organizational challenges also include the need to introduce human oversight mechanisms in line with the requirements of the AI Act. This oversight should not only be formal, but also real – employees must be able to verify the model's decisions and, if necessary, correct them. The European Commission points out that human oversight must include monitoring of results, identification of errors and the possibility of suspending the operation of the system in the event of a threat. This means that banks need to train their staff to interpret model results, understand statistical quality indicators, and analyze the risk associated with algorithmic errors.

One of the biggest challenges is the shortage of specialists who could perform functions related to supervision and algorithmic auditing. As Zetzsche and Buckley (2023) point out, the financial sector has been struggling with a shortage of experts in areas such as artificial intelligence, data engineering, technological ethics or new technology law for years. The implementation of the AI Act means the need to hire new specialists, as well as the expansion of compliance and IT teams. In addition, banks will have to cooperate with external certification bodies and conduct regular audits of models, which generates additional costs.

Another problem is the integration of AI systems with existing banking solutions. Many institutions use legacy systems that are not designed to work with modern learning algorithms. Such integration may require modernization, data migration, and changes in system architecture. The McKinsey report indicates that the lack of compatibility between models and trading systems is one of the biggest barriers to the implementation of advanced AI.

It is also worth paying attention to the cost aspect. The implementation of the AI Act is associated with high costs not only in terms of technology, but also organization, certification, audits and training. Implementation costs may be particularly burdensome for smaller financial institutions that do not have the same resources as the largest European banks (Deloitte, 2025). Therefore, it can be expected that the implementation of the AI Act will lead to further market consolidation and an increase in the importance of large entities capable of investing in advanced technologies.

The implementation of the AI Act in the banking sector is a process that requires extensive changes both on the technological and organizational side. Banks will need to not only adapt their systems and procedures, but also change the way they think about AI, introducing a culture of responsible and transparent model management. The AI Act places high demands on the financial sector, but at the same time it can become an impulse to raise standards of security, quality and trust in banking services. In the long term, the implementation of the new regulations may strengthen the resilience of financial institutions to technological risks and contribute to increasing consumer protection.

CONCLUSIONS

The analysis showed that the Artificial Intelligence Act represents a breakthrough stage in the development of European regulations on artificial intelligence, and plays a particularly important role in the banking sector due to the high technological and systemic risks associated with algorithmic decision-making processes. The research problem – whether the AI Act can realistically increase the security of bank customers – is confirmed by the results of a literature review and comparative analyses. The regulation creates a uniform protective framework that responds to the key risks arising from the use of AI: opacity of models, data bias, algorithmic errors, and the vulnerability of systems to manipulation and cyberattacks.

At the same time, the conclusions indicate that the effectiveness of the AI Act is not guaranteed solely by the regulation itself, but depends on the ability of financial institutions to introduce organizational changes, investments in data infrastructure, and the development of specialized AI governance competencies. Lack of adequate resources, immaturity of model management processes and competency gap can significantly limit the effectiveness of supervision of high-risk systems.

Finally, it should be emphasized that the AI Act is not just a regulatory tool, but a catalyst for the transformation of the entire banking sector. It introduces new standards of technological responsibility, enforces transparency of models and strengthens consumer protection. Institutions that can integrate regulatory requirements with organizational culture and a strategic approach to AI will gain not only legal compliance, but also a lasting competitive advantage based on customer trust and security.

REFERENCES

Barocas, S., Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31

BIS – Bank for International Settlements. (2023). Governance of AI adoption in central banks. https://www.bis.org/publ/othp90.htm

Deloitte (2025). AI and blockchain in financial services, https://www.deloitte.com/us/en/services/audit-assurance/blogs/accounting-finance/ai-blockchain-adoption-in-financial-services.html

EBA – European Banking Authority. (2023). Follow-up report on the use of machine learning for internal ratings-based models. https://www.eba.europa.eu/publications-and-media/press-releases/eba-publishes-follow-report-use-machine-learning-internal

European Commission. (2023). Impact Assessment of the Regulation on Artificial intelligence, https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence

Floridi, L. (2019). Establishing the rules for building trustworthy AI. Nature Machine Intelligence, 1(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y

Goodfellow, I., Shlens, J., Szegedy, C. (2015). Explaining and harnessing adversarial examples. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6572

Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C. R. (2018). Discrimination in the age of algorithms. Journal of Legal Analysis, 10(1), 113–174. https://doi.org/10.1093/jla/laz001

McKinsey & Company. (2023). Global Banking Annual Review 2023: The Great Banking Transition. McKinsey & Company. https://www.mckinsey.com/industries/financial-services/our-insights/global-banking-annual-review-2023

Samek, W., Müller, K. R. (2019). Explainable artificial intelligence: Interpreting, explaining and visualizing deep learning models. Springer. https://doi.org/10.1007/978-3-030-28954-6

Zetzsche, D. A., Buckley, R. P. (2023). The rise of global AI regulation and its impact on financial services. Journal of Financial Regulation, 9(1), 1–28. https://doi.org/10.1093/jfr/fjad005

