


EDITORS

Dimitrios Devetzis, Dana Volosevici,

Leonidas D. Sotiropoulos

A central image showing a grey and black robotic hand shaking a human hand in a dark suit. The hands are clasped in a firm grip. Surrounding the hands are ten yellow five-pointed stars arranged in a circular pattern, reminiscent of the European Union flag. The background is white.

Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order

**Digital Lawscapes:
Artificial Intelligence, Cybersecurity
and the New European Order**

Editors:



Dimitrios DEVETZIS

Activity

Dr. iur. Dimitrios Devetzis (LL.M., M.L.E.), is an Assistant Professor at the Department of Law of Frederick University Cyprus. He completed his LL.B. and first LL.M. on Civil, Civil Procedure and Labour Law, ranked first among all graduates of the respective programs at the Aristotle university of Thessaloniki. He continued his studies in Leibniz Universität Hannover, Germany, where he received his M.L.E. degree in European Private Law (first among all graduates) and finished his PhD with distinction. During his studies he received many state scholarships in Greece (I.K.Y.) and Germany (DAAD). He has worked at the Legal Service of the European Commission and has taught in Greece, Germany, Austria, Czech Republic, Cyprus, Dem. Republic of Congo and other countries. He is affiliated with and hold membership in several prestigious national and international scientific bodies and institutions. In 2024 he was appointed as member of the National Bioethics Committees of the Republic of Cyprus. He is active as Visiting Professor at the Orthodox University of Congo and frequently teaches at Leibniz Universität Hannover. He has participated as a speaker or member of the organizing committee in numerous international conferences and serves as a member or various working groups and advisory bodies regarding with legislative actions. He represents Cyprus at the European and International Private Law Summer School at the Paris Lodron Universität Salzburg. His books have been selected from the libraries of the European Commission and European Court of Justice as a part of their collections. In 2023 he was appointed as Legal Specialist of the European Commission in the field of Civil Law, Law & Technology. He is a Greek Supreme Court (Areios Pagos) appointed attorney and legal counselor of many start-ups and other significant entities. He is fluent in six languages. He serves on the editorial boards of multiple peer-reviewed journals indexed in international databases.

Research Projects and Publications

He is taking part in international research projects regarding contract law – law and economy, as a member of the non profit – research organization *Legal Tech (Athens)*. He is the author or co-author of over 45 academic books and/or articles, among which mention: D. Devetzis, *Daten als Gegenleistung – Einführende Gedanken zu Art 3 der Richtlinie 2019/770/EU* (Göttingen & New York: Vandenhoeck & Ruprecht, 2025 - forthcoming). D. Devetzis, *Die dingliche Surrogation als Rechtsprinzip: Extra legem -Intra ius* (Göttingen: Vandenhoeck & Ruprecht, 2018). D. Devetzis, *The Law of Digital Economy. Legislation – Jurisprudence –*

Commentary & Interpretation (Athens: Nomiki Bibliothiki, 2025). D. Devetzis, *Spouses' Patrimonial Relations during the Period of Separation* (Athens: Nomiki Bibliothiki, 2024). D. Devetzis (2024). *Deep Fake AI and Artists' Employment Contracts under the Lens of the Greek Civil Code*, "ELPIS V-Law Review", 7/2023. D. Devetzis (2024). *Joint Custody and the Principle of the Child's Best Interest in Greek Civil Code after Law 4800/2021*. „Journal of Regional & Socio-Economic Issues”, 14 (2), 6-14. D. Devetzis (2024). *The Role of Private Law in Sustainable Development: Improving Sustainability Through an Effective Analysis of U.S. Contract Law*. „Sustainable Development, Culture & Traditions Journal”, 5 (a), 66-71. D. Devetzis, S. Samaras (2024). *Consumer Protection Safeguards after the AI Act*. „Perspectives of Law and Public Administration”, 13 (2), 298-309. D. Devetzis, (2024). *Smart Contracts and Civil Law: the "modus interpretandi" of contracts "ex machina"*. „Journal of Regional & Socio-Economic Issues”, 14 (3) (36-46). D. Devetzis (2024). *AI, Sustainability Law and EU AI Act*. „Journal of Regional & Socio-Economic Issues”, 14 (3) (18-27). D. Devetzis, S. Samaras (2024). *E-Commerce Platforms and Liability in the AI Era*, „International Investment Law Journal” 4 (1), 18-29.



Dana VOLOSEVICI

Activity

Dr. Dana Volosevici is a Lecturer in Law and Vice Dean for International Cooperation at the Faculty of Letters and Sciences, Petroleum-Gas University of Ploiești, Romania. In parallel with her academic role, she is a practicing attorney with substantial expertise in business law, labor law, data protection, and regulatory compliance. She holds a Bachelor of

Laws from the University of Bucharest and a PhD in Law from the Université de Bretagne Sud, France. Her doctoral thesis — *Analyse multi-facettes de l'intégration des salariés dans les sociétés commerciales, en France et en Roumanie* — was awarded *mention très honorable avec félicitations du jury*, the highest academic distinction granted by the French university system. She has also completed specialization courses at prestigious institutions including the European University Institute - Academy of European Law (Florence), King's College London - Centre of European Law, Cardozo School of Law (New York), Central European University (Budapest), and Hamline University (Minnesota). Dr. Volosevici is an active member of the European Society of International Law (ESIL, Florence), the International Labour and Employment Relations Association (ILERA, Geneva), the Society of Juridical and Administrative Sciences (ADJURIS), and the Association of Privacy and Data Protection Specialists. Her scholarly work frequently addresses issues at the intersection of technology, labor rights, and regulation. Currently, she coordinates the interdisciplinary research project *LS-IN-CLUS 5.0*, which investigates the legal and ethical implications of artificial intelligence in employment contexts, focusing on algorithmic discrimination, inclusion, and governance mechanisms. She is the editor of *Jus et Civitas – Journal of Social and Legal Studies* and serves as a reviewer for several internationally indexed journals in the fields of labor law, technology and regulation, and public policy.

Publications

Dr. Volosevici has published extensively in the areas of labor law, data protection, and legal aspects of digital transformation. Her scholarly output includes numerous articles and book chapters in journals indexed in major international databases such as Web of Science, HeinOnline, DOAJ, and EBSCO. Her most recent article, *Surveillance as a Socio-Technical System: Behavioral Impacts and Self-Regulation in Monitored Environments*, co-authored with Gheorghe Dan Isbășoiu, was published in *Systems* (Q1 Clarivate; IF 2025: 3.1). The article analyzes surveillance through the lens of systems theory, addressing its behavioral effects and implications for legal design in monitored workplaces. She is also the author of *AI Use in the Workplace: Some Legal Risks and Challenges*,

„Analele Universității Alexandru Ioan Cuza din Iași, Seria Științe Juridice”, Vol. 70, No. 2, 2024, pp. 65–77; *The Free Development of Human Personality vs. the Employer’s Right to Monitor the Employee’s Activity in the Era of Generalised Surveillance*, in *Fundamental Constitutional Principles and Their Reflection in the Branches of the Romanian Legal System*, Romanian Academy, 2024, pp. 86–93, ISBN: 978-606-39-1504-8; *Navigating the Complexities of Green Human Resource Management Practices: Operational and Legal Hurdles* (with Gheorghe Dan Isbășoiu), in *Marketing and Resource Management for Green Transitions in Economies*, IGI Global, 2024, pp. 126–153; *Controversial Aspects in the Interpretation of Legal Provisions on Video Surveillance in the Workplace*, *Revista română pentru protecția și securitatea datelor cu caracter personal*, No. 2/2023, pp. 27–36; *Virtual Professional Identity, Legal and Ethical Aspects: A Conceptual Framework* (with Dragoș Grigorescu), *Jus et Civitas*, Vol. X (LXXIV), No. 1/2023, pp. 11–16; *The Digitalisation and the Employment Relationship*, *Journal of Law and Administrative Sciences*, No. 18/2022, pp. 46–52; *Some Considerations on Video Surveillance and Data Protection*, *Jus et Civitas*, Vol. V (LXIX), No. 2/2018, pp. 7–14; *Considerations on the Processing of Personal Data in the Employment Context*, *Economic Insights. Trends and Challenges*, Vol. VI (LXIX), No. 4/2017, pp. 65–72.



Leonidas D. SOTIROPOULOS

Activity

Leonidas D. Sotiropoulos is a Ph.D. candidate at European University Cyprus. His research interests focus on the impact of digitalization and artificial intelligence on marine insurance law. Holding an LL.M. in Shipping Law from Cardiff University, he has extensive experience as a lecturer and certified vocational trainer (HRDA), delivering courses (on-site, on-line) on maritime law, environmental regulations and cyber law for institutions in Greece (Metropolitan College Thessaloniki) and Cyprus. His research focuses on emerging legal and tech challenges in shipping, including autonomous vessels, smart contracts and EU emissions policies, with multiple peer-reviewed publications and conference presentations. With substantial consulting and judicial experience as a partner at a law firm in Thessaloniki, he has practiced in civil, commercial, criminal and administrative law.

Publications

Leonidas Sotiropoulos has established a robust scholarly portfolio at the intersection of maritime law, AI governance, and digital regulation. His 2024-2025 Greek-edition book (printed and e-book) “*Legal Approaches in Maritime Law*” (Hippassus Publishing) provides a comparative analysis of Greek, Cypriot, English, and international maritime and shipping law framework. He contributes two AI-focused chapters to IGI Global’s “*Legal Challenges of AI Across Interdisciplinary Sectors*”, examining smart contract governance and commercial shipping and AI. Recent works include “*The Legal Framework of Autonomous Maritime Technology*” in EVRHYIEL (Ant. Sakkoulas Publishers) and a RAILS-published analysis of the AI Act’s Article 5(1)(f) on emotion recognition (to be published). His maritime and climate policy research features in the *Journal of International Maritime Law* focused on “Decarbonising shipping” and Springer’s “*AI, Sustainability, and Solidarity*” volume, addressing EU ETS and FuelEU Maritime regulations (to be published). Cybersecurity scholarship spans the *Juridical Tribune* (June 2024, *Cyber Challenges amid the Digital Revolution in Maritime Transport*”, paper joint with Professor Kouroupis) and Adjuris conference proceedings (March 2025, *The European Cybersecurity Framework: Challenges, Legal Aspects and Regulations*), while geopolitical analyses like “*Geopolitics and Maritime Terrorism: Challenges and the Legal Complexities in the Red Sea*”(2024) appear in Novi Sad’s proceedings of 21th International Scientific “Legal Days – Prof. Slavko Caric – The Responses of Legal Sciences to The Challenges of modern Society”. Earlier contributions include “*European Maritime Policy and the dynamic of Autonomous Vessels*” (International Investment

Law, 2024), *“Dealing with Smart Contracts in the Insurance Sector: Institutional framework and practical aspects”* (Entha, Cyprus, 2023), *The Legal Treatment of Smart Contracts in English Law*”, (Proceedings of 20th International Scientific “Legal Days – Prof. Slavko Caric-Two Decades of the Development of Legal Thought, Novi Sad, 2023). As editor for ADJURIS’ 2023 business law conference volume, he synthesizes insights on global legal adaptations.

EDITORS

Dimitrios DEVETZIS, Dana VOLOSEVICI,

Leonidas D. SOTIROPOULOS

Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order

Contributions to the 5th International Conference on FinTech,
Cyberspace and Artificial Intelligence Law
March 28, 2025, Bucharest



Bucharest, Paris, Calgary 2025

ADJURIS – International Academic Publisher

This is a Publishing House specializing in the publication of academic books, founded by the *Society of Juridical and Administrative Sciences (Societatea de Stiinte Juridice si Administrative)*, Bucharest.

We publish in English or French treaties, monographs, courses, theses, papers submitted to international conferences and essays. They are chosen according to the contribution which they can bring to the European and international doctrinal debate concerning the questions of Social Sciences.

ADJURIS – International Academic Publisher is included among publishers recognized by **Clarivate Analytics**.

ISBN 978-630-6743-02-5 (E-Book)

© ADJURIS – International Academic Publisher

Editing format .pdf Acrobat Reader

Bucharest, Paris, Calgary 2025

All rights reserved.

www.adjuris.ro

office@adjuris.ro

All parts of this publication are protected by copyright. Any utilization outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution. This applies in particular to reproductions, translations, microfilming, storage and processing in electronic retrieval systems.

Preface

Editors

Assistant professor **Dimitrios Devetzis**,
Frederick University, Cyprus
Lecturer **Dana Volosevici**,
Petroleum-Gas University of Ploiești, Romania
PhD. candidate **Leonidas D. Sotiropoulos**,
European University, Cyprus

This volume contains the scientific papers presented at the 5th International Conference on FinTech, Cyberspace and Artificial Intelligence Law that was held on March 28, 2025, Bucharest, online on Zoom. The conference is organized by the *Society of Juridical and Administrative Sciences* in partnership with the *Charles University in Prague* (ranked 401-500 in the Times Higher Education World University Rankings, 2025 edition). More information about the conference can be found on the official website: https://adjuris.ro/fintech/index_en.html.

The scientific studies included in this volume are grouped into three chapters:

- *Artificial Intelligence: Legal Dimensions, Ethics and Societal Impact.* The papers in this chapter refer to: artificial intelligence - cybersecurity factor; artificial intelligence in employment decision-making: legal challenges and implications; digitalization and AI in anti-corruption efforts: legal challenges, ethical considerations, and future implications.
- *Cybersecurity and Digital Infrastructure Resilience.* This chapter includes papers on: the European cybersecurity framework: challenges, legal aspects and regulations; NIS2 Directive - legal preparedness of EU health infrastructure against large-scale cyberattacks; bank digitalization and virtual agents driving financial inclusion and data protection.
- *Regulation, Data Protection and AI Governance in the EU.* The papers in this chapter refer to: the new EU Product Liability Directive - interaction with parallel EU initiatives: proposed AI Liability Directive, Digital Services Act and Digital Markets Act; digital rights in the age of artificial intelligence: challenges and perspectives; Artificial Intelligence Act and GDPR: implications for AI solution developers and users in Romania; the use of artificial intelligence in combating tax evasion: challenges, opportunities and ethical implications from a legal perspective.

The articles included in this volume have been revised using the "double blind" peer review system, respecting international scientific standards.

This volume is aimed at practitioners, researchers, students and PhD candidates in cyberspace and artificial intelligence law, who are interested in recent developments and prospects for development in this field at international and national level.

We thank all contributors and partners and are confident that this volume will meet the needs for growing documentation and information of readers in the context of globalization and the rise of dynamic elements in AI law.

Table of Contents

ARTIFICIAL INTELLIGENCE: LEGAL DIMENSIONS, ETHICS AND SOCIETAL IMPACT	14
<i>Adriana-Iuliana STANCU</i>	
Artificial Intelligence, Cybersecurity Factor.....	15
<i>Dana VOLOSEVICI</i>	
Artificial Intelligence in Employment Decision-Making: Legal Challenges and Implications	25
<i>Mădălina VOICAN</i>	
Digitalization and AI in Anti-corruption Efforts: Legal Challenges, Ethical Considerations, and Future Implications	42
CYBERSECURITY AND DIGITAL INFRASTRUCTURE RESILIENCE	56
<i>Leonidas SOTIROPOULOS</i>	
The European Cybersecurity Framework: Challenges, Legal Aspects and Regulations	57
<i>Antonia RENGLE</i>	
NIS2 Directive - Legal Preparedness of EU Health Infrastructure Against Large-Scale Cyberattacks	72
<i>Isabelle OPREA, Daniela DUȚĂ</i>	
Bank Digitalization and Virtual Agents Driving Financial Inclusion and Data Protection	92
REGULATION, DATA PROTECTION AND AI GOVERNANCE IN THE EU.....	110
<i>Dimitrios DEVETZIS</i>	
The New EU Product Liability Directive. Interaction with Parallel EU Initiatives: Proposed AI Liability Directive, Digital Services Act and Digital Markets Act	111

Aurel Octavian PASAT

Digital Rights in the Age of Artificial Intelligence: Challenges and Perspectives	144
--	-----

*Camelia Daciana STOIAN, Dominic BUCERZAN, Radu Nicolae STOIAN,
Catalin Raul HALIC, Crina Anina BEJAN*

Artificial Intelligence Act and GDPR: Implications for AI Solution Developers and Users in Romania	162
---	-----

Mihai ȘTEFĂNOAIA

The Use of Artificial Intelligence in Combating Tax Evasion: Challenges, Opportunities, and Ethical Implications from a Legal Perspective	178
---	-----

**ARTIFICIAL INTELLIGENCE:
LEGAL DIMENSIONS, ETHICS AND
SOCIETAL IMPACT**

Artificial Intelligence, Cybersecurity Factor

Associate professor **Adriana-Iuliana STANCU**¹

Abstract

Objectives: Recent cyberattacks on significant European institutions, the exponential rise in cyberthreats, and the speed at which technology is developing have brought attention to the need for increased cooperation and change in the civil-military sphere and the fact that there is no hierarchy between the military and civilian communities. As mandated by international agreements, including those pertaining to the Charter, the EU's cybersecurity policy enables it and its Member States to improve their ability to defend, detect, protect, and even prevent by appropriately utilizing the entire spectrum of security options at the civilian and military communities. Proposals and Methodology: The need to defend European values and invest in their preservation has led to the EU's cooperation structures becoming involved in the cyber offensive, including with its financial capabilities, even though each EU member state has direct responsibility for its national security, including in the sensitive cyber domain, as a direct result of Article 4(2) TEU. Results and Implications: To defend the EU, its citizens, the EUIBA, and their operations and missions in the cyber domain related to the Permanent Security and Defence Policies (PSDP), it is imperative that the actions of all European nations and European institutions, organizations, and agencies, including EUIBA, be strengthened in the upcoming period. Additionally, it highlights the need of cyber resilience at the EU level by boosting defensive capabilities in this delicate, cutting-edge area, expanding the potential for cyber defence, and generating trustworthy input from Member States. Thus, cooperation is required to improve cybersecurity.

Keywords: EU member state; cybersecurity; European values; cyber domain.

JEL Classification: K14

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/01>

Please cite this article as:

Stancu, Adriana-Iuliana, „Artificial Intelligence, Cybersecurity Factor”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 15-24.

1. Introduction

The technological development and innovations that humanity has devel-

¹ Adriana-Iuliana Stancu - “Dunarea de Jos” University of Galati, Romania, ORCID: 0000-0001-6259-5116, adriana.tudorache@ugal.ro.

oped in the recent years have reached unimaginable dimensions, and thus the development social and economic process in the technological, medical, cultural and military area in the last half century has taken different forms, whether we are talking about the defense industry, the car manufacturing industry or the IT field, which in some cases are interdependent. When the first microprocessor was invented in the early 1970s, information technology became widely used, and processing speed increased dramatically. Shortly after that, internet networks were widely used by the public, and the number of computers connected reached several hundred. This number increased even more, eventually connecting many people and accelerating the development of blockchain technology and quantum computing.

The management of data, information, and knowledge and their transformation into an optimal best management practice (BMP) for a positive outcome are linked to the evolution of computerization and computing equipment. As a result, in addition to the growth in data volume, their storage capacity has also increased, moving from the byte of the 1960s to the yottabyte and brontobyte of the modern era. The increase in storage capacity and the amount of data has required the subsequent development of computing tools, which, from classical tools and traditional computers, has reached quantum computing and the development of quantum computers that generate a real-time computing method and, in some cases, surpassing the understanding capacity of the human mind; however, the introduction of new technologies has frequently had a profound impact on society.

2. Cybersecurity in the European Union

The Council has decided on a common approach to EU cybersecurity policy regarding the 2014 cyber defence modality and its 2018 amendments. This includes reinvesting in our modern, cooperative forces, technologies, and next-generation capabilities, as well as cybersecurity and fortifying partnerships to tackle shared challenges².

Informatics has become an area of strategic competition at a sensitive time due to the increasing use of digital technology. Thus, it is necessary to maintain a permanent online presence, free from external influences, secure and unchanged. Information technology has facilitated Russia's war of aggression against Ukraine, which has impacted the entire world and contributed to instability and insecurity with a significant risk of permanent escalation³. It has also generated more internet activity than this senseless and brutal conflict.

² Forbrukerradet, *Deceived by Design. How tech companies use dark patterns to discourage us from exercising our rights to privacy*, 2018, p. 6, <https://storage02.forbrukerradet.no/media/2018/06/2018-06-27-deceived-by-design-final.pdf>, accessed on 15 March 2025.

³ Jonna Järveläinen, Duong Dang, Mike Mekkanen, and Tero Vartiainen. 2025. "Towards a Framework for Improving Cyber Security Resilience of Critical Infrastructure against Cyber Threats: A

The Russian-initiated war in Ukraine has created a new strategic context and shown why European nations, the Union as a whole, and its allies must further solidify the EU's stance to eradicate cyberthreats and to bolster traditional cybersecurity and cyber defences against criminal activity and cyber security attempts in the "online" sphere.

The European institutions' will to respond quickly and effectively to threats that aim to compromise, interfere with, or take control of networks and IT systems, among other things, is emphasized in the Joint Communication on the EU Cybersecurity Policy. This Joint Communication represents a new accomplishment in the EU's comprehensive approach to resilience, response, conflict prevention, connectivity, and stability in the single cyberspace by updating the Cybersecurity Strategy and taking it to the global level in accordance with the strategic guidelines. In this context, Member State representatives stressed the need for appropriate and consistent responses from EU, its Member States and its partners, who are on standby, to the review of the guidelines for the implementation of the EU Cyber Diplomacy Toolkit as a new step development of cyber platform⁴.

In applying the provisions of the 2014 Cyber Defense Policy Framework and the abdication of the next 4 years, Member State representatives, the Council, agreed on a common approach on EU cybersecurity policy to reinvest in our modern and cooperative forces and technologies and next-generation capabilities, as well as cybersecurity and strengthening partnerships to solve common problems⁵.

The cyber development area has become an area of strategic competition at a time when dependence on digital technologies is increasing. Thus, it is necessary to maintain an open, independent, stable and secure online presence. The use of these computers that sparked and followed Russia's unprovoked and still wholly unjustified war of aggression against Ukraine threatens international stability and security, poses a serious risk of escalation, and adds to the already notable rise in Internet activity that occurs outside of the recent armed conflict.

From a strategic perspective, the war in Ukraine is a novel situation that has once again demonstrated the necessity for the EU, its member states, and its partners to continue supporting the EU in creating solutions to cybercrime and to uphold its reputation for cybersecurity and defence against criminal activity and

Dynamic Capabilities Approach." *Journal of Decision Systems* 34 (1). doi: 10.1080/12460125.2025.2479546. Also see R. Srinivasan, M. Kavitha, R. Kavitha, and S. Uma (2023). "Cybersecurity and Artificial Intelligence: A Systematic Literature Review." In Sugumaran D, Souvik Pal, Dac-Nhuong Le, Noor Zaman Jhanjhi (eds.), *Recent Trends in Computational Intelligence and Its Application*. Proceedings of the 1st International Conference on Recent Trends in Information Technology and its Application (ICRTITA, 22) 1st ed., CRC Press, London, p. 120 et seq. <https://doi.org/10.1201/9781003388913>.

⁴ Lilian Edwards, Michael Veale (2017), „Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For”. 16 *Duke Law & Technology Review*, pp. 18-84, p. 67.

⁵ Forbrukerradet, *op. cit.*, p. 6.

tendentious acts on the “online” space. This collaboration demonstrates a new step toward a comprehensive EU vision on resilience, feedback, the eradication of conflict ideas, cooperation, and stability in the cyber space. It fulfils its cybersecurity thinking once strategic concepts are unified. In this situation, the Council of Europe has shown that direct and well-coordinated responses are needed from the EU area, its Member States and its partners, who, in turn, strongly want to review the possibilities of implementing all the instruments for cyber diplomacy in the EU as a step up in the development of this cyber platform⁶.

A step-by-step, visible and penetrating approach is essential for the development of trust, which in turn is necessary for the future establishment of a crisis management structure in the EU and beyond, in terms of constructive stability in relation to cybersecurity in this generous space. The plan thus conceived regarding crisis management is being developed by the Council. It also resumes and discusses the need to continue to develop our capabilities to defend, detect, defend and stop criminal cyber-attacks through a significant penetration of the area of knowledge of what is happening, capacity building, capacity development, training, testing and a special resistance as a non-returnable response against cyber-attacks directed towards European countries and EUIBA, the missions in the CSDP theatre of operations, using all existing possibilities. In doing so, the Council supports the High Representative and the Commission to control cyberspace, not to get involved in the management of pointless work and to ensure collaboration with existing initiatives. Understanding and coordination of European countries cybersecurity professionals must be supported in a planned manner, between all communities, both military and civilian, in the online space and between a public and a private ecosystem that inspires trust. In this situation, Member States are supported to research and permanently develop national mechanisms for civil-military cooperation, thus facilitating the mutual exchange of information, collaborate on lessons learned, contribute to supporting interoperable standards and create risk assessments by building reliable platforms for man-made or natural disasters, as well as cooperative operations, in particular at European level but also with other states, in full compliance with the European legislative provisions on the measures required to strengthen its development, exceptional cybersecurity⁷.

Thus, online education, training and exercises are put back in the foreground, as they are essential to ensure their availability and effectiveness, but also because new jobs are needed at national level, through services originating from the EU space through the European Security and Defence Academy (ESDA), EDA, ENISA and the future introduction of PESCO, through projects such as the

⁶ Lilian Edwards, Michael Veale, *op. cit.*, p. 67.

⁷ Carolina Polito, Lorenzo Pupillo (2024), „Artificial Intelligence and Cybersecurity”, *Forum Journal*, Volume 59, No. 1, p. 10-13. For a few connections with human rights see Rowena Rodrigues, “Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities.” *Journal of Responsible Technology* 4 (December 2020): 100005. <https://doi.org/10.1016/j.jrt.2020.100005>.

Internet Environment Associations and the EU Internet Academy as well as the Innovation Hub (CAIH). But to consolidate these efforts, the European institutions are concerned with the establishment of the EDA CyDef-X framework project to coordinate and support cybersecurity services. The Council is responsible for the development of the EDA – European Defence Agency to investigate, in close cooperation with European countries with EEAS⁸, how CyDef-X can also support activities such as CYBER PHHALANX, including mutual support in compliance with the provisions of Article 42, paragraph 7 of the TEU but also in accordance with the solidarity clause as clearly follows from the provisions of Article 222 of the TFEU, as well as the Commission and ENISA in relation to civil actions. Furthermore, the Council supports the use of the CyDef-X cybersecurity test area through continuous development. Today, there are also beneficial proposals such as Cyber Range Federations. To ensure a rapid and efficient decision-making process regarding an unresolved situation in a cyber crisis, the Council points out that it is necessary to permanently organize exercises at national and mass level in the decision-making matter of the Member States.

3. Cybersecurity and Artificial Intelligence

The presence of unacceptable risks posed with AI used in ways not permitted by law will inevitably lead to prohibitions and general provisions of the 2025 Use Regulation. Although the overall effectiveness of each prohibition is linked to the establishment of control and. In the application of this Regulation, the intended use of prohibitions is essential to explain the risks that may cause disasters and to be reflected effectively in other processes such as civil law. Furthermore, the most important infrastructure management system and policy consideration must be functional before 2 August 2026, when the legislative provision will be in field. Thus, what refers to notified bodies but also to the governance structure must be applicable from 2 August 2025. In the first era of technological growth but also the adoption AI models with the clear aim of general use, the roles of AI model providers should be aligned with the general application from 2 August. The AI Office must consider that classification policies and practices are updated and comply with new technological developments. In view of all this, Member States must establish and inform the European institutions about rules sanctions, reconsider whether they are applicable when decree in question will entry into force⁹.

These harmonization provisions set out in the Regulation should apply to all sectors and, subject to the new legal framework, existing Union legislation

⁸ European Union External Action Service.

⁹ Daron Acemoglu, *Opinion: The AI we should fear is already here*, in The Washington Post (2021), in <https://www.washingtonpost.com/opinions/2021/07/21/ai-we-should-fear-is-already-here/>, accessed on 15 March 2025.

should not be affected, in particular the GDPR protection which are already guaranteed, of operational workers, but also product safety, which complement the Regulation¹⁰, the compensation amount for any damages incurred, as stipulated in Council Directive 85/374/EEC, is still in effect and completely enforceable. Furthermore, the Regulation is opposed by its provisions to Union law in relation to social policy and legal provisions in labor law, relating to employment and the protection of workers, to working conditions in general, fair practices in the field of employment, safety at work, including cooperation between employers and employees. In a positive sense, the Regulation does not call into question the existence of fundamental rights in democratic exercise in the Member States of the Union, including the right correlated with the freedom to know and to carry out other types of activities associated with certain operational systems belonging to members of the governments of European countries, such as the right to mediation, to conclude collective agreements and to apply them or to take joint action, in compliance with European legal provisions for by national law¹¹.

Those providing physical or virtual components must bear in mind As AI systems created for direct action with people have been created, developed and approved so that people interact directly with the AI system, with one exception when the matter is obvious for well-founded reasons, which a prudent or just attentive person, in relation to the existing circumstances and the level at which it is used, can establish through his own perceptions. The obligation does not lie with AI systems authorized by law to investigate, determine, prevent or pursue the commission of crimes in the field, in relation to the protection of the rights and freedoms of third parties, given that such systems exist in the public domain for the detection of crimes.

Individuals or legal entities that market AI systems, including AI systems that create artificial content, i.e. transform voices, images, which can be video or text, must be sure that the results of the AI system can be processed, to establish the artificial creation or use. Those who create such systems must consider that the solutions are efficient, credible but also with collaborative potential and with operational possibilities depending on the content, costs but also on the development of the technology that must be public, according to the standards in force. The obligation is not applicable when AI systems perform a routine editing auxiliary function or do not modify the substance of the data provided by the implementer or its content or if they are required by law to detect, prevent, investigate or detect crime¹².

¹⁰ <https://www.gov.uk/data-protection>, accessed on 15 March 2025.

¹¹ Daniel J. Solove, *The digital person. Technology and Privacy in the Information Age*, New York University Press, 2004, p. 22 et seq., https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2501&context=faculty_publications, accessed on 15 March 2025.

¹² Working Party, "Article 29" *On Data Protection, Guidelines on automated individual decision-making and profiling under Regulation (EU) 2016/679*, Adopted on 3 October 2017 as last revised and adopted on 6 February 2018, p. 13, https://ec.europa.eu/newsroom/document.cfm?doc_id=47742, accessed on 15 March 2025.

Where these providers belong to third countries, they must send a representative to the European Union to specify the requirements of the Order.

Regarding persons who implement emotional recognition systems or use biometric classification, they are obliged by law (informed consent) to inform the individuals thus exposed about the functioning of the system and its use for the processing of personal data, as set out in the Regulations in the field promoted at European level as well as in the 2016 Directive. The obligation presented does not apply to AI systems created specifically for biometric classification and establishing emotions, used in the criminal field and under the aegis of criminal laws, for the prevention and investigation of crimes, in compliance with the provisions of the Code of Criminal Procedure created in compliance with the rights and freedoms recognized to all persons, because they are under the aegis of European legislation¹³.

Those implementing the AI system that creates or uses images, voices or video fragments develops deepfakes that indicate artificially composed or used content. This obligation is not applicable if its use is in the spirit of the provisions of criminal law for the detection, prevention, prosecution or repression of crimes¹⁴. When this content is part of a work or program of a known artistic, creative, satirical, opinion-based or similar nature, the obligations to disseminate information contained in the legal provision in question, have the sole recognized purpose of highlighting the processed or manipulated content but which does not contradict the exposition or agreement of the work. The requirements thus requested become mandatory when the entire EU legal system cannot be undermined¹⁵.

Implementers of an AI system generate, or process published documents publicly displayed for the purpose of informing the public about situations important to individuals indicating that this book was written or used. The obligation itself does not apply when the use is provided for and accepted according to the law to establish, investigate or prosecute “detectable crimes or if the content created by AI was investigated through an editorial sample responsibility for publication” or and. content belonging to a natural or legal person.

Information must be delivered to those interested as clearly and precisely as possible, up until the moment of the first presentation or contact. This information must be established according to accessibility requirements.

Thus, the AI office supports and creates the possibility of developing and promoting good practices policies throughout the Union to facilitate the efficient

¹³ Lilian Edwards, Michael Veale, *op. cit.*, p. 67; Forbrukerradet, *op. cit.*, p. 6.

¹⁴ Adriana Iuliana Stancu, (2024). „Combating The Financing Terrorism: an Analysis of the EU Regulatory Framework and Enforcement Mechanism”, in Ojars Sparitis (ed.), *Proceedings of 11th SWS International Scientific Conference on Social Sciences - ISCSS 2024*, SGEM WORLD SCIENCE (SWS) Scholarly Society, DOI: 10.35603/sws.iscss.2024/s02/06.

¹⁵ Agencia Española de Protección de Datos (2020), *RGPD compliance of processing that embed Artificial Intelligence. An introduction*, 2020, p. 6, <https://www.aepd.es/guides/gdpr-compliance-processings-that-embed-ia.pdf>, accessed on 15 March 2025.

establishment of functions in relation to the determination and identity of substances produced or created by this inventive process. The Commission may adopt any legislation to implement or approve each of the elements of good practice¹⁶.

The moment when the supervisory authority of a European country establishes with clear evidence that an AI system presents elements of risk, according to the Regulation, an assessment of the performance possibilities of the AI system is required, in relation to the requirements and obligations requested that go beyond the framework of the Regulation, with an emphasis on AI systems with special risk in relation to vulnerable groups of people. The supervisory authority is only required to notify and work in permanent cooperation with the government that is in charge of the controls or with the pertinent entities mentioned in the Order when dangers to fundamental rights are identified.

4. Conclusions

Beyond the Framework Decision's content, the EU's examination of cybercrime must be viewed in the context of its significant importance. The scope of cybercrime issues covered by this thorough investigation is substantially wider, which directly contributes to the identification of legislative tools of significant relevance to the EU that impact both the first and third pillars of the EU. The development of the areas of freedom, justice, and security is where the third pillar's battle against cybercrime lies.

Making a broader criticism, namely in the light of the fact that the Framework Decision was verified by the European Commission in close connection with the case C-176/03 of the Court of Justice of the EU, regarding the division of competences in the field of criminal cases, reported to the European Commission and the Council of the EU, a permanent reflection is required on the issues raised by the legislation on cybercrime, at the EU institutional level. Another issue that requires increased attention is the way in which the EU solves the thorny issue of data protection, in parallel with the legislation at the global level that approaches this topic from a different perspective, creating a negative impact. However, we must acknowledge that the EU has added value in combating cybercrime in the areas of freedom, justice, and security that define it. The Council of Europe Treaty on Cybercrime is especially valuable in this regard. It transcends the boundaries of the EU Framework Decision and is distinguished by the fact that any nation interested in resolving this complex issue can do so. This new feature in the field of Council instruments has been used for the first time. The ratification of the Framework Decision was included, relatively recently, in the annex to the Communication from the Commission to the European Parliament

¹⁶ Anuța Gianina Opre, Simona Șandru, „The right to be forgotten on the internet, a means of combating discrimination” in M. Tomescu (ed.), *Non-discrimination and equal opportunities in contemporary society*, Pro Universitaria Publishing House, Bucharest, 2015, p. 288.

and the Council on the results of the 2005 Court judgment in case C176/03.

All the States signatories to the Treaty have embarked on a common path, with a sustained effort, and with an extended range of action, possibly at global level, in the active fight against cybercrime.

A consensus is desired on the measures required to control cybercrime and of course a total elimination of cybercrime, but these are unlikely goals to be achieved, and so the cyberspace will always have a space to fight to make things right. What is wonderful is that remarkable progress has been made in this almost unknown area and common solutions have been found to address cybercrime. Not every issue has been resolved or identified to date, but coordinating efforts to develop a more comprehensive definition of cybercrime and standardizing laws in this innovative area across all EU member states are crucial steps in the European fight against this new type of highly intrusive, cross-border crime.

Bibliography

1. Acemoglu, Daron, *Opinion: The AI we should fear is already here*, in The Washington Post (2021), in <https://www.washingtonpost.com/opinions/2021/07/21/ai-we-should-fear-is-already-here/>, accessed on 15 March 2025.
2. Agencia Española de Protección de Datos (2020), *RGD compliance of processing that embed Artificial Intelligence. An introduction*, 2020, <https://www.aepd.es/guides/gdpr-compliance-processings-that-embed-ia.pdf>, accessed on 15 March 2025.
3. Edwards, Lilian & Michael Veale (2017), „*Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For*”. 16 *Duke Law & Technology Review*, pp. 18-84.
4. Forbrukerradet, *Deceived by Design. How tech companies use dark patterns to discourage us from exercising our rights to privacy*, 2018, <https://storage02.forbrukerradet.no/media/2018/06/2018-06-27-deceived-by-design-final.pdf>, accessed on 15 March 2025.
5. Järveläinen, Jonna, Duong Dang, Mike Mekkanen, and Tero Vartiainen. 2025. “Towards a Framework for Improving Cyber Security Resilience of Critical Infrastructure against Cyber Threats: A Dynamic Capabilities Approach.” *Journal of Decision Systems* 34 (1). doi: 10.1080/12460125.2025.2479546.
6. Opre, Ancuța Gianina & Simona Șandru, „The right to be forgotten on the internet, a means of combating discrimination” in M. Tomescu (ed.), *Non-discrimination and equal opportunities in contemporary society*, Pro Universitaria Publishing House, Bucharest, 2015.
7. Polito, Carolina & Lorenzo Pupillo (2024), „Artificial Intelligence and Cybersecurity”, *Forum Journal*, Volume 59, No. 1, p. 10-13.
8. Rodrigues, Rowena, “Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities.” *Journal of Responsible Technology* 4 (December 2020): 100005. <https://doi.org/10.1016/j.jrt.2020.100005>.
9. Solove, Daniel J., *The digital person. Technology and Privacy in the Information Age*, New York University Press, 2004, https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2501&context=faculty_publications, accessed on

15 March 2025.

10. Srinivasan, R., M. Kavitha, R. Kavitha, and S. Uma (2023). "Cybersecurity and Artificial Intelligence: A Systematic Literature Review." In Sugumaran D, Souvik Pal, Dac-Nhuong Le, Noor Zaman Jhanjhi (eds.), *Recent Trends in Computational Intelligence and Its Application*. Proceedings of the 1st International Conference on Recent Trends in Information Technology and its Application (ICRTITA, 22) 1st ed., CRC Press, London, <https://doi.org/10.1201/9781003388913>.
11. Stancu, Adriana Iuliana (2024). „Combating The Financing Terrorism: an Analysis of the EU Regulatory Framework and Enforcement Mechanism”, in Ojars Sparitis (ed.), *Proceedings of 11th SWS International Scientific Conference on Social Sciences - ISCSSL 2024*, SGEM WORLD SCIENCE (SWS) Scholarly Society, DOI: 10.35603/sws.iscss.2024/s02/06.
12. Working Party, "Article 29" *On Data Protection, Guidelines on automated individual decision-making and profiling under Regulation (EU) 2016/679*, Adopted on 3 October 2017 as last revised and adopted on 6 February 2018, https://ec.europa.eu/newsroom/document.cfm?doc_id=47742, accessed on 15 March 2025.

Artificial Intelligence in Employment Decision-Making: Legal Challenges and Implications

Lecturer **Dana VOLOSEVICI**¹

Abstract

This study explores the legal challenges arising from the use of artificial intelligence in employment decision-making, with particular attention to its implications for employees' rights and managerial accountability. The primary objective is to evaluate whether existing legal frameworks, most notably the General Data Protection Regulation and the Artificial Intelligence Act, offer sufficient protection against the risks associated with automated and algorithmically informed decisions in the workplace. Employing a qualitative methodology, the research is grounded in doctrinal analysis of relevant European legal instruments, supplemented by a review of academic literature in labour law, data protection, and algorithmic governance. The study adopts an interdisciplinary perspective, combining legal analysis with insights from organisational psychology and data science. The findings underscore key concerns, including the risk of indirect discrimination, the opacity of algorithmic decision-making processes, and the potential dilution of managerial responsibility. In response, the paper recommends a series of organisational measures such as targeted training, structured collective bargaining on AI deployment, and the adoption of a sustainable, rights-oriented approach to managing the workforce. The study concludes that a multidimensional governance model is essential to ensure that technological innovation remains aligned with the protection of workers' fundamental rights and the principles of democratic workplace governance.

Keywords: artificial intelligence; employment decision-making; labour law; data protection; employee participation.

JEL Classification: K24, K31

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/02>

Please cite this article as:

Volosevici, Dana, „Artificial Intelligence in Employment Decision-Making: Legal Challenges and Implications”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 25-41.

1. Introduction

Artificial Intelligence has become increasingly embedded in employment

¹ Dana Volosevici - Petroleum Gas University of Ploiesti, Romania, ORCID: 0009-0000-9109-4679, dana.volosevici@upg-ploiesti.ro.

decision-making processes due to its potential to enhance efficiency, objectivity, and predictive accuracy. In contemporary labour markets, employers face complex demands to evaluate vast amounts of candidate and employee data swiftly and consistently. AI systems offer algorithmic solutions that can process such data in ways that human decision-makers may find time-consuming or prone to cognitive bias. As noted in legal scholarship, this technological shift corresponds with a broader move toward data-driven governance in employment practices, which is reshaping traditional managerial prerogatives².

The integration of AI into employment settings also involves the continuous and large-scale processing of personal data, using a huge capacity of processing. This means that the volume, granularity, and frequency of data collection significantly exceed traditional methods of employee supervision, enabling employers to draw complex inferences about individual behaviour, productivity, and even psychological traits. The scale and sophistication of such data processing raise important concerns regarding the proportionality and necessity of surveillance practices, particularly when viewed through the lens of data protection principles enshrined in Article 5 of the General Data Protection Regulation³.

Moreover, AI monitoring may be extended beyond standard working hours. For example, remote work monitoring software may continue to track device usage or location data after official working time has ended, and algorithmic scheduling tools may access personal calendars or communications to optimise future workflows. This blurs the boundary between professional and private life, potentially infringing upon the worker's right to disconnect and to a private life under Article 8 of the European Convention on Human Rights and Article 7 of the Charter of Fundamental Rights of the European Union. In this regard, the continuous reach of AI tools into employees' time and personal space not only challenges traditional notions of managerial control but also necessitates a recalibration of labour law protections in light of emerging technologies.

Against this background, the legal framework has had to evolve to address the risks of hyper-surveillance and to ensure that digital monitoring does not erode the dignity and autonomy of workers. Initially, the GDPR established a comprehensive framework aimed at safeguarding personal data and reinforcing fundamental rights in the digital age. Building upon this foundation, AI Act introduced specific obligations for high-risk AI systems, including those used in employment contexts, thereby ensuring that the development and deployment of

² Valerio De Stefano, *Negotiating the Algorithm: Automation, Artificial Intelligence and Labour Protection*, „Comparative Labor Law & Policy Journal” 41(1), 2019, p. 15–46.

³ Paul De Hert & Vagelis Papakonstantinou, *The new General Data Protection Regulation: Still a sound system for the protection of individuals?* „Computer Law & Security Review”, 32(2), 2016, p. 179–194; Antonie Aloisi & Elena Gramano, *Artificial Intelligence is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context*, „Comparative Labor Law & Policy Journal”, 41(1), 2019, p. 95–126.

AI technologies respect the fundamental rights and freedoms of the human person.

More specifically, GDPR introduced a series of foundational principles designed to enhance individuals' control over their personal data and to ensure that data processing activities are carried out lawfully, fairly, and transparently. Among these, the principle of transparency — enshrined in Article 5(1)(a) GDPR — requires that data subjects be clearly informed about how their personal data is collected, used, and for what purposes. Transparency is further reinforced by Articles 12 to 14 GDPR, which impose obligations on data controllers to provide information in a concise, intelligible, and easily accessible form, using clear and plain language. This principle is especially crucial in the context of automated decision-making, where individuals may otherwise lack insight into how algorithmic processes influence outcomes that affect them directly⁴. In addition to transparency, the GDPR codifies the principles of data minimisation and purpose limitation (Article 5(1)(c) and (b), respectively), which require that only personal data that is adequate, relevant, and limited to what is necessary in relation to the purposes for which it is processed be collected and used.

Crucially, Article 22 GDPR establishes the right not to be subject to a decision based solely on automated processing, including profiling, where such a decision produces legal effects or significantly affects the individual. This provision reflects a broader concern about protecting human dignity and autonomy in the face of algorithmic decision-making. The right under Article 22 is particularly relevant in employment contexts, where automated decisions about hiring, task allocation, or performance evaluation may profoundly impact workers' professional and personal lives.

The AI Act classifies as high-risk AI systems both those intended to be used for the recruitment or selection of natural persons — particularly for the targeted placement of job advertisements, the analysis and filtering of job applications, and the evaluation of candidates — and those intended to be used for making decisions affecting the terms of employment relationships. This includes decisions relating to promotion, termination of contractual employment relationships, the allocation of tasks based on individual behaviour or personal traits or characteristics, as well as the monitoring and evaluation of the performance and conduct of individuals engaged in such relationships.

As a consequence, such AI systems must comply with a series of requirements expressly laid down in Articles 9 to 15 of the AI Act, including the obligation to ensure human oversight. This requirement, as defined in Article 14, entails the implementation of appropriate human involvement during the operation of the AI system, with the aim of preventing or minimising risks to health, safety,

⁴ Sandra Wachter, Brent Mittelstadt, Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, „International Data Privacy Law”, Volume 7, Issue 2, May 2017, p. 76–99, <https://doi.org/10.1093/idpl/ixp005>.

or fundamental rights. Human oversight must be effective, meaning that individuals designated to supervise the system are able to understand its functioning, detect anomalies or inappropriate outcomes, and intervene when necessary to override or disable the system's operations.

The purpose of this study is to examine whether, despite the existing legislative provisions, risks related to the use of AI systems in employment-related decision-making still persist. The analysis will focus on three main aspects. First, it will explore the nature and specificity of managerial decision-making when such decisions are informed or influenced by the outcomes generated by AI systems. Secondly, the study will examine the operational characteristics of AI systems, with particular attention to those aspects of their functioning that may impact employees' rights. Finally, it will seek to identify potential organisational measures that could be implemented to mitigate the risk of infringing employees' rights, thereby promoting the responsible use of AI in the workplace.

2. The Specificity of AI-Informed Managerial Decisions in Employment

The increasing integration of artificial intelligence into human resources management has begun to reshape the nature and legitimacy of managerial decision-making. While AI systems promise efficiency, standardisation, and data-driven objectivity, their use in employment contexts raises complex questions about human agency, legal accountability, and the boundaries of managerial discretion.

a) Human Versus AI Reasoning in Employment Decision-Making. One of the foundational issues that merits legal and interdisciplinary examination is the distinction between human reasoning and AI-based reasoning in employment-related decisions. Traditionally, employment relations have evolved through decisions made by human agents, who increasingly rely on sophisticated data analytics and predictive tools. However, the decisional logic of human actors remains qualitatively different from that of algorithmic systems. Human managerial decisions are typically shaped by a complex interplay of experience, contextual judgment, organisational values, and often tacit knowledge accumulated over time⁵. Although such decisions may be partially subjective, they also allow for adaptability to specific contexts, such as the organisational climate, the socio-economic environment, or the individual characteristics of workers. In this sense, human discretion accommodates the flexibility needed to uphold key legal and ethical values in labour relations, including fairness, proportionality, and individualisation.

⁵ Michael J.R. Butler, Holly L.R. O'Broin, Nick Lee and Carl Senior, *How Organisational Cognitive Neuroscience Can Deepen Understanding of Managerial Decision-Making: A Review of the Recent Literature and Future Directions*, „International Journal of Management Reviews“, Vol. 18, 2016, p. 542–559.

By contrast, AI systems reason through computational logic, deriving statistical correlations from historical data and optimising decisions based on pre-defined objectives, such as efficiency or productivity. These systems do not interpret context or moral values; they function by identifying patterns and applying learned associations, absent a normative framework⁶. Consequently, while AI-generated outcomes may appear more precise or objective, they lack the interpretive flexibility that human reasoning provides—particularly relevant in labour law, where many decisions require case-by-case assessment⁷. This divergence creates a structural asymmetry in the decision-making process. Human reasoning allows for ethical reflection and legal proportionality, while AI reasoning, even when technically robust, may unintentionally reproduce biases embedded in training data⁸. For instance, recruitment algorithms trained on past hiring decisions may internalise and perpetuate historical patterns of gender or racial discrimination, even if unintentionally⁹.

At a legal level, this distinction becomes particularly salient. Employment decisions often entail significant consequences, such as promotion, disciplinary sanctions, or termination, and must comply with substantive and procedural labour standards. These include the right to equal treatment, due process, and protection against unfair dismissal, as recognised in both national labour laws and EU law (Article 30 of the Charter of Fundamental Rights of the European Union; Council Directive 2000/78/EC). The application of generalised algorithmic logic to such decisions can undermine the required individualised assessment, risking violations of anti-discrimination law or procedural safeguards.

Therefore, the use of algorithmic management tools in the workplace carries the risk of creating a false perception of neutrality, whereby decisions appear objective but are, in fact, removed from the legal and ethical principles that have traditionally underpinned human-centred employment relationships. This disconnect is further intensified by the inherent limitations of AI systems, which are unable to account for non-quantifiable factors such as empathy, moral reasoning,

⁶ Adams-Prassl, Jeremias, *What if Your Boss Was an Algorithm? The Rise of Artificial Intelligence at Work*, „Comparative Labor Law & Policy Journal”, Vol. 41(1), 2019, Available at SSRN: <https://ssrn.com/abstract=3661151>.

⁷ Valerio De Stefano and Mathias Wouters, *AI and Digital Tools in Workplace Management Evaluation: An Assessment of the EU's Legal Framework*, European Parliamentary Research Services, Scientific Foresight Unit (PE 729.516), 2022.

⁸ Sara Baiocco, Enrique Fernández-Macías, Uma Rani, and Annarosa Pesole, *The Algorithmic Management of Work and its Implications in Different Contexts*, European Commission, 2022; Katherine C Kellogg, Melissa A Valentine and Angèle Christin, *Algorithms at Work: The New Contested Terrain of Control*, „Academy of Management Annals”, Vol. 14, 2020, p. 366; Mohammad Hossein Jarrahi, *Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organisational Decision Making*, „Business Horizons”, Vol. 61(4), 2018, p. 577–586 <https://doi.org/10.1016/j.bushor.2018.03.007>.

⁹ Miriam Kullmann, *Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law*, „International Journal of Comparative Labour Law and Industrial Relations”, Vol 34(1), 2018, p. 1–21.

or evolving organisational norms, elements that remain essential for fair and lawful decision-making in human resource management.

b) *Automation Bias and Managerial Deference.* Another layer of complexity in AI-informed managerial decision-making stems from documented psychological tendencies that shape how humans interact with automated systems. A central concept in this regard is automation bias - the cognitive inclination to trust and over-rely on decisions or suggestions made by automated systems, even in the face of contradictory or incorrect outputs. Research has shown that individuals tend to defer to automated recommendations because these systems are perceived as faster, more accurate, or more neutral than human judgment¹⁰. This bias is particularly pronounced in high-stakes or high-pressure environments, such as employment decision-making, where managerial accountability, time constraints, and organisational expectations converge. In such contexts, managers may perceive AI systems as more reliable, consistent, or legally defensible than their own subjective judgment, especially when the outputs are presented with high confidence or a veneer of objectivity.

From a legal and organisational perspective, this dynamic has significant implications. Article 22 of the GDPR grants individuals the right not to be subject to decisions based solely on automated processing, including profiling, that produces legal effects or significantly affects them. However, this right assumes the existence of genuine and informed human oversight. If managerial review is reduced to a mere formal endorsement, driven by automation bias and limited understanding, then the safeguard becomes functionally ineffective. Similarly, Article 14 of the AI Act requires that high-risk AI systems be subject to effective human oversight. However, the regulation does not explicitly define the cognitive or organisational preconditions needed for oversight to be meaningful. Research suggests that effective oversight is not merely procedural but requires a critical mental model of how the system works, as well as the confidence and autonomy to override its outputs when necessary¹¹.

c) *Responsibility and Accountability in AI-Informed Decisions.* Despite

¹⁰ Goddard K, Roudsari A, Wyatt JC, *Automation bias: a systematic review of frequency, effect mediators, and mitigators*, „Journal of the American Medical Informatics Association”, 19(1), 2012, p.121-127. doi: 10.1136/amiajnl-2011-000089; Hannah Ruschemeier and Lukas J. Hondrich, *Automation Bias in Public Administration – An Interdisciplinary Perspective from Law and Psychology*, „Government Information Quarterly”, 41(3), 2024, 101953, <https://doi.org/10.1016/j.giq.2024.101953>.

¹¹ Linda J. Skitka, Kathleen Mosier and Mark D. Burdick, *Accountability and Automation Bias*, „International Journal of Human-Computer Studies”, 52(4), 2000, p. 701–717, <https://doi.org/10.1006/ijhc.1999.0349>; Eugenio Alberdi, Lorenzo Strigini, Andrey A. Povyakalo, Peter Ayton, *Why Are People's Decisions Sometimes Worse with Computer Support?*. In: Buth, B., Rabe, G., Seyfarth, T. (eds) „Computer Safety, Reliability, and Security”. SAFECOMP 2009. Lecture Notes in Computer Science, vol 5775. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04468-7_3.

the growing reliance on AI systems in employment-related decision-making, legal responsibility remains with human actors, typically the employer and, in some instances, individual managers. Under established principles of employment law, decision-makers must be able to justify their decisions —whether those decisions are adverse, such as disciplinary action or dismissal, or favourable, such as promotion or the award of a performance bonus. The obligation to justify decisions stems from the requirement to demonstrate their legality, fairness, and proportionality, particularly in contexts where such decisions may significantly affect the employee's rights and professional trajectory.

However, when AI systems, especially those based on machine learning and characterised by limited interpretability, are involved in the decision-making process, managers may find themselves unable to explain the rationale behind a particular recommendation. The technical complexity and opacity of such systems, often referred to as "black box" models, mean that key elements of the decision may originate from a source external to the manager's control, yet the manager remains compelled, or feels pressured, to rely on those outputs.

This situation raises important questions regarding the allocation of legal responsibility. In other regulatory domains, such as product liability, one might consider the role of the AI system developer or manufacturer. However, in the field of labour law, responsibility for employment decisions continues to rest with the employer, including responsibility for the choice and implementation of AI tools used within the organisation. Given the inherently asymmetrical nature of the employment relationship, where the employer holds structural and legal power over the employee, it is the employer who must bear responsibility for organising work, including decisions informed by AI¹². Accordingly, even when decisions are mediated or informed by algorithmic systems, the notion of a "responsibility gap"¹³, where no actor is held accountable, is not admissible within the framework of labour law. The employer must ensure that the systems deployed in human resource management are compatible with legal standards and that adequate mechanisms for oversight and contestability are in place.

Nevertheless, when decisions are based on highly complex AI systems, even where these systems meet the formal requirements of the AI Act, employers may face significant challenges in demonstrating that such decisions are proportionate, lawful, and non-discriminatory. This difficulty may undermine the employer's ability to defend its actions before labour courts or data protection authorities, especially in jurisdictions where the burden of proof in discrimination cases lies with the employer.

¹² Valerio De Stefano and Simon Taes, *Algorithmic Management and Collective Bargaining*, „Transfer: European Review of Labour and Research”, 29(1), 2022, <https://doi.org/10.1177/10242589221141055>.

¹³ Andreas Matthias, *The responsibility gap: Ascribing responsibility for the actions of learning automata*, „Ethics and Information Technology”, 6, 2004, p. 175–183. <https://doi.org/10.1007/s10676-004-3422-1>.

Furthermore, the diffusion of decision-making agency between algorithmic systems and human actors fundamentally challenges the principle of personal accountability that underlies managerial functions. Employment law has traditionally operated on the assumption that decision-makers act with deliberation, autonomy, and legal awareness. When managers are unable to interrogate, override, or meaningfully interpret algorithmic outputs, this assumption no longer holds. As a result, the integrity of decision-making processes, and the legal accountability attached to them, risks being substantially weakened, unless clear governance frameworks and robust safeguards are instituted to preserve human oversight and legal responsibility.

3. Algorithmic Design and the Protection of Workers' Rights

A central concern in the governance of artificial intelligence is the distinction between opaque and transparent AI systems, a distinction with significant implications for the protection of fundamental rights, particularly in the employment context. These categories refer not only to the technical architecture of AI models but also to their legal relevance in determining accountability, explainability, and compliance with data protection and non-discrimination obligations.

Opaque AI systems, often referred to as “black box” models, are characterised by their limited interpretability. Such systems, typically based on advanced machine learning techniques such as neural networks, generate outputs without providing a comprehensible rationale that could be understood by non-expert users, including employers or affected individuals. This opacity inhibits the ability to interrogate, justify, or challenge decisions, thereby undermining core principles of transparency, accountability, and procedural fairness¹⁴. Conversely, transparent AI systems are either inherently interpretable, such as those employing decision trees or rule-based logic or are supplemented by explainability tools that enable users to understand the logic behind outputs. Transparent systems are more easily aligned with legal standards that require decisions to be explainable, contestable, and reviewable, particularly when they have significant effects on individuals, as is often the case in employment. The AI Act directly addresses these concerns by imposing a series of obligations on high-risk AI systems. Under Article 13 AI Act, providers of high-risk AI systems must ensure that the system is designed and developed in a manner that allows for an appropriate level of transparency. This includes the obligation to inform users about the system’s intended purpose, its decision-making logic (where possible), and any limitations that may affect its reliability or fairness.

However, even in the case of transparent AI systems, the decision-mak-

¹⁴ Amedeo Santosuosso and Giovanni Sartor (2024), *Decidere con l’IA: Intelligenze artificiali e naturali nel diritto*, Il Mulino, p. 75.

ing process remains inherently complex. It typically involves multiple stages, including data collection, data preprocessing, algorithmic analysis, and the generation of recommendations. These phases, which are not under the direct control of the system user, must ensure accuracy, relevance, and legal compliance from the system's initial design phase and throughout its operational lifecycle¹⁵.

For example, the data collection phase involves the systematic aggregation of relevant data from multiple sources, both within the organisation and external to it. These sources include structured data, such as employee attendance records, performance metrics, and financial transactions, as well as unstructured data, such as emails, feedback reports, or written evaluations. AI-driven recruitment platforms, for example, utilize resumes, online job applications, interview transcripts, and previous hiring decisions to assess candidate suitability¹⁶. Similarly, compliance monitoring systems integrate data from internal policy documents, regulatory frameworks, and past legal rulings to evaluate employees' adherence to workplace regulations. A significant aspect of data collection may involve real-time data streaming, where AI continuously receives updates from workplace monitoring tools, biometric attendance systems, customer feedback surveys, and other dynamic sources¹⁷.

For algorithmic analysis to be statistically valid and unbiased, it must process a large and representative dataset that includes heterogeneous sources and temporally distributed data points. The inclusion of varied sources ensures that the algorithm captures diverse patterns and avoids biases introduced by homogeneous data. Processing data from different time periods accounts for temporal shifts, preventing models from becoming outdated or overfitting to short-term trends. These factors are critical for ensuring algorithmic fairness, predictive accuracy, and robustness in dynamic workplace environments¹⁸. For statisticians, data must be valid and reliable, meaning that the input data should have a verified level of accuracy to reduce the likelihood of Type I errors (false positives). To obtain such data, it is mandatory to ensure rigorous data collection processes.

As data complexity increases, so does the risk of violating legal provisions designed to protect the rights of data subjects, particularly employees in this context. This situation creates an inherent tension between two legitimate interests: statisticians, who require large-scale, detailed data to develop statistically

¹⁵ Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, Patrick Vink, *Fair, Transparent, and Accountable Algorithmic Decision-making Processes*, „Philosophy and Technology”, 31, 2018, p. 611–627. <https://doi.org/10.1007/s13347-017-0279-x>; Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi, *The ethics of algorithms: Mapping the debate*, „Big Data & Society”, 3(2), 2016, <https://doi.org/10.1177/2053951716679679>.

¹⁶ Aaron Rieke and Miranda Bogen, *Help wanted: An examination of hiring algorithms, equity, and bias*, Labor and Employment. Upturn, 2018, December 10. <https://www.upturn.org/reports/2018/hiring-algorithms/>, accessed 16 March 2025.

¹⁷ Ifeoma Ajunwa, *The quantified worker: Law and technology in the modern workplace* Cambridge University Press, 2023. <https://doi.org/10.1017/9781316888681>.

¹⁸ Sandra Wachter, Brent Mittelstadt, Luciano Floridi, *op.cit.*

robust models, and legal professionals, who work to ensure that data processing remains transparent, proportional, and compliant with privacy and data protection regulations. While statisticians prioritize data completeness and granularity to enhance the accuracy and predictive power of their models, legal experts emphasize the need to minimize data collection, limit processing to strictly necessary information, and ensure fairness and accountability in its use.

The data preprocessing phase, which transforms raw data into a usable format, could also raise legal concerns. Data cleaning, which involves correcting errors or filling in missing values, may (unintentionally) alter employee records, raising issues of data integrity and legal accountability, especially in sensitive areas as disciplinary or performance-related disputes. Normalization and standardization, used to harmonize data formats, can facilitate accurate analysis but may inadvertently perpetuate systemic biases, especially if contextual differences (e.g., regional pay scales) are not properly accounted for. In feature selection and engineering, the identification of variables for algorithmic decision-making must be carefully scrutinized. Including or proxying protected characteristics (e.g., gender or ethnicity) risks indirect discrimination, unless justified under specific legal frameworks such as affirmative action policies. Employers must ensure that these design choices comply with both anti-discrimination law and the AI Act's data governance requirements (Article 10 AI Act). However, fulfilling these obligations presupposes that employers possess sufficient technical and legal knowledge to identify, request, and verify compliance with such requirements. This represents a significant shift in the traditional scope of managerial responsibility, requiring not only legal awareness but also a degree of algorithmic literacy. Moreover, trade unions and employee representatives should also be equipped to engage with these systems meaningfully, particularly in their roles concerning consultation, co-determination, and the protection of workers' rights.

Once data has been preprocessed, AI systems apply algorithmic models to identify patterns and classify behaviours. While such techniques are intended to increase the efficiency and consistency of managerial decision-making, they also introduce significant legal risks, particularly in relation to discrimination, data protection, and procedural fairness. A brief review of several common AI techniques helps to illustrate the potential risks at stake. Predictive analytics, used to anticipate outcomes such as employee attrition or performance decline, typically relies on historical datasets. However, these datasets may reflect past inequalities or biased decision-making, thereby embedding and perpetuating indirect discrimination against protected groups¹⁹. Classification models, which assign employees or behaviours into predefined categories (e.g., "low risk," "high risk"), may produce disparate impacts when seemingly neutral input variables correlate

¹⁹ Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishing, 2016.

with protected characteristics such as age, gender, or ethnicity. This raises concerns regarding algorithmic discrimination, which may infringe the principles of equal treatment and proportionality in employment law.

Anomaly detection systems, often employed in compliance or fraud monitoring, identify deviations from normative behaviour. However, where such models are poorly calibrated or based on incomplete contextual information, they may generate false positives. The resulting actions — such as unwarranted disciplinary investigations — may infringe on workers' reputational rights, procedural guarantees, and the broader principle of fair treatment under both labour law and fundamental rights frameworks.

Following the algorithmic analysis phase, AI systems generate outputs in various forms — such as recommendations, rankings, or alerts — tailored to the specific decision-making context in which they are deployed. While these outputs are intended to enhance managerial efficiency and consistency, they raise significant concerns regarding interpretability, contextual relevance, and legal accountability. These concerns arise primarily from the fact that AI-generated outputs, though often perceived as objective or neutral, may in fact be opaque, offering limited insight into the reasoning behind specific recommendations — thereby increasing the risk of discriminatory outcomes²⁰, particularly when the underlying data or model design reflects existing societal or institutional biases.

As previously noted, the transition from algorithmic output to human intervention is not without legal risks. Human decision-makers may adjust AI-generated outputs, override recommendations, or request additional information before reaching a final decision. However, this presupposes that the human agent treats the AI system as a standard decision-support tool, engaging with it through critical reasoning rather than passive acceptance. For such oversight to be meaningful, the human operator must possess the requisite technical and legal knowledge to interpret the system's output and must be provided with sufficient information by the AI system itself — such as the underlying rationale, confidence levels, and limitations — to enable an informed and lawful decision.

4. Internal Governance of Workplace AI Systems

As the deployment of AI systems in employment contexts becomes increasingly prevalent, the need for organisational safeguards to prevent the infringement of employees' rights has become a pressing legal and ethical concern. While regulatory frameworks such as the GDPR and AI Act provide general obligations concerning data governance, transparency, and human oversight, these must be complemented by internal organisational measures to ensure compliance and accountability at the operational level.

²⁰ Jeremias Adams-Prassl, Reuben Binns and Aislinn Kelly-Lyth, *Directly Discriminatory Algorithms*, *Modern Law Review*, 86(1), 2023, p. 144-175. <https://doi.org/10.1111/1468-2230.12759>.

a) *AI Literacy in the Workplace.* A first and essential organisational measure to ensure the responsible use of AI in the workplace is the education and training of both managers and employees on the structure, functioning, and legal implications of the AI systems in use. This is particularly important given the increasing complexity and opacity of algorithmic decision-making tools, which often function as black boxes and produce outputs that may appear objective but are not easily interpretable.

Managers, in particular, must develop a basic level of AI literacy — the ability to understand how algorithmic systems operate, the types of data they use, the logic of their decision-making processes, and the contexts in which their outputs are reliable or problematic. This knowledge is critical not only for effective oversight but also for complying with legal obligations, such as those laid down in Article 14 of the AI Act, which requires human oversight mechanisms for high-risk systems. Without sufficient training, managers may over-rely on AI outputs, treating them as authoritative even in cases where critical contextual judgment is required.

Workers, likewise, must be informed about the presence and role of AI in decisions affecting their working lives, including areas such as recruitment, task allocation, performance evaluation, and disciplinary action. The GDPR reinforces this obligation through its principles of transparency (Article 5(1)(a)) and the right to be informed (Articles 13 and 14). Workers should be made aware of what data is being collected, how it is processed, and for what purposes it is used. This transparency is not merely procedural but essential to enable workers to contest adverse decisions, exercise their right to explanation (under Article 15 GDPR), and assert their rights under Article 22 GDPR, which prohibits decisions based solely on automated processing that produce significant effects unless specific safeguards are in place.

Moreover, training initiatives should not be limited to isolated sessions or technical documentation. Instead, they should be integrated into broader organisational learning strategies and include interactive elements — such as case studies, simulations, and critical discussions — that foster a culture of accountability and ethical awareness. As argued by De Stefano and Taes²¹, embedding these capacities within both managerial practice and workforce participation is essential to prevent the erosion of labour rights in algorithmically managed environments.

b) *Worker Participation in the Regulation of Algorithmic Management.* A second, and equally essential, organisational safeguard is the integration of collective bargaining and social dialogue into the governance of AI systems used in employment contexts²². The deployment of algorithmic tools in the workplace should not be left to unilateral managerial discretion, particularly given

²¹ Valerio De Stefano and Simon Taes, *op.cit.*

²² *Idem.*

their capacity to affect fundamental rights such as privacy, non-discrimination, and the right to fair working conditions. As emphasised by the European Trade Union Confederation²³ and recognised in broader EU labour law frameworks, worker participation through collective bargaining is a key mechanism to ensure democratic oversight over technological change.

Collective agreements can and should serve as regulatory instruments capable of defining the scope, purpose, and limitations of AI use in the workplace. This includes stipulating the types of decisions for which AI tools may be used, setting clear boundaries around invasive monitoring practices, and determining when human oversight must intervene. By formalising these parameters, collective bargaining helps to operationalise the principles of proportionality and necessity, which underpin both the GDPR and the AI Act.

Moreover, collective agreements can establish procedures for transparency and auditing, particularly for high-risk systems used in areas such as recruitment, task allocation, disciplinary measures, and performance evaluation. These procedural safeguards are essential for ensuring compliance with Article 22 GDPR, which prohibits decisions based solely on automated processing unless specific conditions and safeguards are met, including the right to obtain human intervention and to contest the decision. In this context, algorithmic auditing — an internal or external review of how the AI system functions, how data is processed, and whether outputs are fair and legally compliant — can be embedded into collective agreements as a recurring obligation, not merely a one-time procedural formality.

c) *A Risk-Based Approach to AI Governance.* Organisations should adopt a risk-based approach to the deployment of AI in the workplace, recognising that not all processes require algorithmic optimisation. Employers ought to limit the use of AI systems to contexts where their benefits are demonstrably proportionate to the potential legal, ethical, and social risks. This approach aligns with the principle of data minimisation under Article 5(1)(c) GDPR, as well as the risk management framework established by Article 9 of the AI Act, which requires providers of high-risk systems to implement continuous, iterative processes to identify, evaluate, and mitigate risks.

This cautious deployment strategy is also consistent with the objectives of Sustainable Development Goal 8 (SDG 8), which promotes decent work and inclusive economic growth. In particular, SDG 8 calls for the protection of labour rights and the promotion of safe and secure working environments for all workers. The indiscriminate or opaque use of AI systems — particularly in hiring, monitoring, or disciplinary decisions — risks undermining these commitments by eroding transparency, fairness, and accountability in the employment relationship. If employees are to be recognised as legitimate stakeholders whose rights

²³ European Trade-Union Confederation (ETUC), *ETUC Resolution Calling for an EU Directive on Algorithmic Systems at Work*, 3 December 2021, <https://www.etuc.org/en/document/etuc-resolution-calling-eu-directive-algorithmic-systems-work> accessed 17 March 2025.

and dignity are respected within organisational structures, then ethical employment practices must include a responsible approach to the use of AI. Given that the AI Act explicitly classifies employment-related AI systems as high-risk (Annex III, Section 4), it is incumbent upon employers to demonstrate that their use of such systems is not only lawful but also aligned with broader principles of corporate social responsibility and sustainable workplace governance.

To institutionalise ethical and legal oversight, organisations should adopt an internal code of conduct governing the use of AI in employment. Such a code should clearly articulate key principles — transparency, fairness, non-discrimination, and human oversight — and set out internal procedures and responsibilities for ensuring compliance. Beyond policy statements, organisations may establish AI ethics committees or appoint AI compliance officers tasked with reviewing proposed AI systems, evaluating their impact on workers' rights, and overseeing their ongoing use. These internal governance mechanisms can serve not only to mitigate legal liability but also to reinforce organisational accountability and cultivate trust among employees.

5. Discussions

The findings of this paper contribute to the evolving discourse on how artificial intelligence reshapes the nature of decision-making in employment, particularly by influencing or informing managerial judgments. While AI has introduced unprecedented capabilities in terms of data processing, pattern recognition, and predictive analytics, its deployment in employment contexts, especially in high-stakes areas such as recruitment, performance evaluation, and disciplinary decisions, raises complex legal challenges.

At the heart of the debate is the epistemological divergence between human and algorithmic reasoning. Human managers operate within a normative framework that values fairness, proportionality, and contextual judgment, as embedded in both national labour law and EU legal standards. By contrast, AI systems optimise outputs based on statistical correlations in training data, which may be shaped by historical patterns of discrimination or systemic inequalities. This divergence is particularly problematic when algorithmic outputs are perceived as neutral or objective, despite being generated through opaque processes that lack moral or legal reasoning. As a result, there is a risk that AI systems could normalise and obscure discriminatory practices, while simultaneously disempowering human decision-makers who defer to them.

This leads to a second area of concern: the risk of automation bias and the erosion of managerial accountability. Users frequently place undue trust in algorithmic systems, especially when the systems are marketed as tools for enhancing efficiency and fairness. In practice, however, this reliance may result in rubber-stamping AI outputs without proper scrutiny. The AI Act's requirement for effective human oversight (Article 14) and the GDPR's Article 22 safeguards

both assume a level of AI literacy and critical engagement that is not guaranteed in most workplace settings. If decision-makers do not understand how a system functions, or cannot interrogate the logic behind its outputs, they cannot fulfil their legal duties, nor can they meaningfully justify their decisions in the event of litigation.

Another key insight emerging from this study is the essential role of collective bargaining and worker participation in the governance of AI systems in the workplace. As emphasised by the European Trade Union Confederation (ETUC, 2021), the adoption of algorithmic management tools should not occur through unilateral employer decisions. Rather, social dialogue must serve as the framework for democratically negotiating the scope, conditions, and safeguards of AI deployment. One of the most critical areas for negotiation concerns the limitation of AI use, particularly in light of the significant volume of data processing that even transparent AI systems require. The functioning of such systems often entails the collection and analysis of extensive employee-related data, which, in certain cases, results in excessive or disproportionate monitoring practices.

Although AI systems rely on access to large and organisation-specific datasets to deliver relevant outputs, this technical requirement cannot justify the continuous surveillance of workers. The protection of privacy and dignity in the workplace must remain a core principle, and any AI deployment must be scrutinised to ensure that it does not place employees under pervasive or intrusive monitoring regimes.

Furthermore, the discussion has placed the responsible use of AI within the broader context of sustainable and ethical workplace governance. As suggested by the risk-based approach in the AI Act (Article 9) and the principle of data minimisation under the GDPR (Article 5(1)(c)), organisations should not pursue algorithmic optimisation indiscriminately. Instead, they must assess whether AI use is proportionate to the task at hand and whether the legal, ethical, and social risks outweigh the operational benefits. This approach also resonates with Sustainable Development Goal 8 (SDG 8), which advocates for decent work, labour protections, and inclusive economic growth. AI deployment that undermines workers' autonomy, privacy, or procedural rights is incompatible with these broader commitments.

Finally, the discussion affirms that regulation alone is not sufficient. While the GDPR and AI Act establish key principles and obligations, their effectiveness depends on the extent to which organisations internalise and operationalise these norms. This requires not only legal compliance but also a shift in organisational culture, one that values human agency, transparency, and responsibility in the face of increasingly automated infrastructures.

In sum, the findings suggest that a multidimensional governance strategy is needed, one that integrates legal safeguards, technical design, organisational

practices, and participatory mechanisms. Only through such an integrated approach can the transformative potential of AI be harnessed in a way that upholds the fundamental rights and dignity of workers.

Bibliography

1. Adams-Prassl, Jeremias, Binns, Reuben and Kelly-Lyth, Aislinn, *Directly Discriminatory Algorithms*, “Modern Law Review”, 86(1), 2023, pp. 144–175. <https://doi.org/10.1111/1468-2230.12759>.
2. Adams-Prassl, Jeremias, *What if Your Boss Was an Algorithm? The Rise of Artificial Intelligence at Work*, “Comparative Labor Law & Policy Journal”, Vol. 41(1), 2019. Available at SSRN: <https://ssrn.com/abstract=3661151>.
3. Ajunwa, Ifeoma, *The Quantified Worker: Law and Technology in the Modern Workplace*, Cambridge University Press, 2023. <https://doi.org/10.1017/9781316888681>.
4. Alberdi, Eugenio, Strigini, Lorenzo, Povyakalo, Andrey A. and Ayton, Peter, *Why Are People’s Decisions Sometimes Worse with Computer Support?*, in Buth, B., Rabe, G. and Seyfarth, T. (eds), “Computer Safety, Reliability, and Security”, SAFECOMP 2009, Lecture Notes in Computer Science, vol. 5775, Springer, Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-642-04468-7_3.
5. Aloisi, Antonio and Gramano, Elena, *Artificial Intelligence is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context*, “Comparative Labor Law & Policy Journal”, 41(1), 2019, pp. 95–126.
6. Baiocco, Sara, Fernández-Macías, Enrique, Rani, Uma and Pesole, Annarosa, *The Algorithmic Management of Work and Its Implications in Different Contexts*, European Commission, 2022.
7. Butler, Michael J.R., O’Broin, Holly L.R., Lee, Nick and Senior, Carl, *How Organisational Cognitive Neuroscience Can Deepen Understanding of Managerial Decision-Making: A Review of the Recent Literature and Future Directions*, “International Journal of Management Reviews”, Vol. 18, 2016, pp. 542–559.
8. De Hert, Paul and Papakonstantinou, Vagelis, *The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals?*, “Computer Law & Security Review”, 32(2), 2016, pp. 179–194.
9. De Stefano, Valerio and Taes, Simon, *Algorithmic Management and Collective Bargaining*, “Transfer: European Review of Labour and Research”, 29(1), 2022. <https://doi.org/10.1177/10242589221141055>.
10. De Stefano, Valerio and Wouters, Mathias, *AI and Digital Tools in Workplace Management Evaluation: An Assessment of the EU’s Legal Framework*, European Parliamentary Research Services, Scientific Foresight Unit (PE 729.516), 2022.
11. De Stefano, Valerio, *Negotiating the Algorithm: Automation, Artificial Intelligence and Labour Protection*, “Comparative Labor Law & Policy Journal”, 41(1), 2019, pp. 15–46.
12. Goddard, Kate, Roudsari, Ali and Wyatt, Jeremy C., *Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators*, “Journal of the

- American Medical Informatics Association”, 19(1), 2012, pp. 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>.
13. Jarrahi, Mohammad Hossein, *Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organisational Decision Making*, “Business Horizons”, 61(4), 2018, pp. 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>.
14. Kellogg, Katherine C., Valentine, Melissa A. and Christin, Angèle, *Algorithms at Work: The New Contested Terrain of Control*, “Academy of Management Annals”, Vol. 14, 2020, p. 366.
15. Kullmann, Miriam, *Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law*, “International Journal of Comparative Labour Law and Industrial Relations”, 34(1), 2018, pp. 1–21.
16. Lepri, Bruno, Oliver, Nuria, Letouzé, Emmanuel, Pentland, Alex and Vink, Patrick, *Fair, Transparent, and Accountable Algorithmic Decision-Making Processes*, “Philosophy and Technology”, 31, 2018, pp. 611–627. <https://doi.org/10.1007/s13347-017-0279-x>.
17. Matthias, Andreas, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, “Ethics and Information Technology”, 6, 2004, pp. 175–183. <https://doi.org/10.1007/s10676-004-3422-1>.
18. Mittelstadt, Brent Daniel, Allo, Patrick, Taddeo, Mariarosaria, Wachter, Sandra and Floridi, Luciano, *The Ethics of Algorithms: Mapping the Debate*, “Big Data & Society”, 3(2), 2016. <https://doi.org/10.1177/2053951716679679>.
19. O’Neil, Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishing, 2016.
20. Rieke, Aaron and Bogen, Miranda, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, Labor and Employment, Upturn, 10 December 2018. <https://www.upturn.org/reports/2018/hiring-algorithms/> accessed 16 March 2025.
21. Ruschemeier, Hannah and Hondrich, Lukas J., *Automation Bias in Public Administration – An Interdisciplinary Perspective from Law and Psychology*, “Government Information Quarterly”, 41(3), 2024, 101953. <https://doi.org/10.1016/j.giq.2024.101953>.
22. Santosuosso, Amedeo and Sartor, Giovanni, *Decidere con l’IA: Intelligenze Artificiali e Naturali nel Diritto*, Il Mulino, 2024.
23. Skitka, Linda J., Mosier, Kathleen and Burdick, Mark D., *Accountability and Automation Bias*, “International Journal of Human-Computer Studies”, 52(4), 2000, pp. 701–717. <https://doi.org/10.1006/ijhc.1999.0349>.
24. Wachter, Sandra, Mittelstadt, Brent and Floridi, Luciano, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, “International Data Privacy Law”, 7(2), 2017, pp. 76–99. <https://doi.org/10.1093/idpl/ix005>.

Digitalization and AI in Anti-corruption Efforts: Legal Challenges, Ethical Considerations, and Future Implications

Associate professor **Mădălina VOICAN**¹

Abstract

This research explores the role of Digitalization and Artificial Intelligence (AI) in detecting, preventing, and predicting corruption, addressing both their potential and the legal challenges they present. Traditional anti-corruption mechanisms rely heavily on human intervention, yet they often suffer from inefficiency, limited adaptability. Meanwhile, AI-driven technologies have emerged as powerful tools for enhancing fraud detection, financial monitoring, and procurement oversight. The study further examines how the new technologies expand these capabilities by enabling predictive analytics to anticipate corruption risks before they materialize, offering a more proactive approach to combating corruption. However, the deployment of AI in anti-corruption efforts raises legal and ethical concerns, particularly regarding the black-box nature of AI models, algorithmic bias, and transparency. To mitigate these risks, this study discusses the importance of accountability and regulatory enforcement, emphasizing the need for robust legal frameworks, clear regulatory standards, and ethical guidelines for AI implementation. The research concludes that while AI has the potential to revolutionize anti-corruption efforts, its success depends on strong legal safeguards and responsible governance.

Keywords: artificial intelligence, digitalization, anti-corruption, legal frameworks, algorithmic bias, transparency, governance.

JEL Classification: K14, K24

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/03>

Please cite this article as:

Voican, Madalina, „Digitalization and AI in Anti-corruption Efforts: Legal Challenges, Ethical Considerations, and Future Implications”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 42-55.

1. Context: Why Fighting Corruption Matters

Corruption remains a serious global challenge that significantly holds back economic growth, erodes investor confidence, and diminishes national

¹ Mădălina Voican - Faculty of Law, University of Craiova, Romania, ORCID: 0000-0003-0876-0437, madalina.voican@edu.ucv.ro.

productivity in many countries². This negative impact on economic growth is further supported by World Bank studies, which suggest that when countries make progress in reducing corruption (as shown by better scores on corruption indexes), they often experience stronger or faster economic growth at the same time. Which means that, improvements in measures of corruption and growth accelerations tend to go hand-in-hand as the two trends support each other³. Therefore, less corruption usually leads to better use of public money, more trust from investors, and more efficient institutions, all of which help the economy grow. On the other hand, when the economy grows quickly, countries have more resources and capacity to improve governance and reduce corruption.

Therefore, as a persistent societal illness corruption continues to cause harm around the world by weakening public institutions, eroding trust in government, limiting access to essential services, and deepening social and economic inequalities⁴. It diverts public resources away from development priorities, fuels injustice, and often leaves the most vulnerable members of society without protection or opportunity⁵.

According to the 2024 Corruption Perceptions Index (CPI) published by Transparency International, more than two-thirds of the 180 countries assessed scored below 50 out of 100, indicating serious levels of public sector corruption. More precisely, 118 countries scored below 50, which represents approximately 66% of all countries evaluated. These results highlight the persistence of corruption worldwide and emphasize the continued need for strong efforts to combat this issue⁶.

Recognizing the systemic and destructive nature of corruption, the United Nations Convention Against Corruption (UNCAC, 2005) has provided a comprehensive framework and set of recommendations to guide Member States in their efforts to prevent and combat corruption effectively.

In response to the systemic and pervasive harm resulting from corruption, and considering its serious consequences for countries' development, modern

² Per Aarvik, 2019. *Artificial Intelligence – a promising anti-corruption tool in development settings?*, Bergen: CMI - CHR Michelsen Institute, (U4 Report 2019:1), p. 23, <https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings>, accessed on 27.05.2025.

³ World Bank Group Flagship Report (2025), *Global Economic Prospects*, International Bank for Reconstruction and Development/The World Bank, Washington, DC, p. 175-180, <https://openknowledge.worldbank.org/server/api/core/bitstreams/f983c12d-d43c-4e41-997e-252ec6b87dbd/content>, accessed on 27.05.2025.

⁴ Ousmane Diagana & Mouhamadou Diagne (2023). *La corruption est un problème mondial pour le développement. Pour la combattre, nous avons tous un rôle à jouer*. La Tribune Afrique, June 2023, <https://www.banquemoniale.org/fr/news/opinion/2023/06/13/corruption-is-a-global-problem-for-development-to-fight-it-we-all-have-a-role-to-play>, accessed on 27.05.2025.

⁵ Myint, U., 2000, *Corruption: Causes, consequences and cures*. Asia-Pacific Development Journal, 7(2), pp. 33-57.

⁶ Transparency International. (2024). *Corruption Perceptions Index 2024*. <https://www.transparency.org/en/cpi/2024/index/aze>, accessed on 27.05.2025.

technology must be part of the solution to combat it effectively, as researchers demonstrate that anti-corruption policies aimed at reducing discretion and enhancing institutional quality are significantly strengthened by digital tools.

Many experts underline that by significantly reducing human intervention, new technologies minimize opportunities for corruption⁷. This reduction in human intervention improves accuracy and reduces costs associated with manual processing and human error, making processes more efficient. Several studies have explored the application of AI in combating corruption⁸ showing how AI has been utilized to detect fraudulent activities, analyze financial transactions, and monitor procurement processes more than ever. Moreover, the use of AI highlights its potential as powerful tools for identifying and preventing corruption, as they can process and analyze complex datasets much faster and more accurately than traditional methods, making them valuable assets in anti-corruption strategies⁹.

2. Methodology of Research

The study aims to explore the theoretical foundations of Digitalization and AI its application in detecting, preventing, and potentially predicting corruption. By examining the unique capabilities of recent technologies, which can generate content and identify patterns without human-imposed criteria, this research seeks to understand the relationship between these advanced models and anti-corruption measures.

The methodology involves a comprehensive literature review, conceptual analysis, theoretical modeling, and the evaluation of ethical, legal, and social implications.

The study is guided by the research question of how Digitalization and Artificial Intelligence (AI) can be theoretically and practically applied to detect prevent and predict corruption within a framework considering ethical, legal and social implications.

3. Digitalization Role in Anti-corruption Strategies

Corruption is a complex phenomenon that is difficult to measure directly, as it usually happens in secret. People involved in corruption try to hide their

⁷ Nils Köbis, Jean-François Bonnefon. & Iyad Rahwan (2021), „Bad machines corrupt good morals”. *Nature Human Behaviour* 5, 679–685. <https://doi.org/10.1038/s41562-021-01128-2>.

⁸ Félix J. López-Iturriaga, Iván Pastor Sanz (2018), „Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces”. *Social Indicators Research* 140, 975–998. <https://doi.org/10.1007/s11205-017-1802-2>; Fernanda Odilla (2023) „Bots against corruption: Exploring the benefits and limitations of AI-based anti-corruption technology”, *Crime, Law and Social Change Journal*, Volume 80(4), pp. 1-44, <https://doi.org/10.1007/s10611-023-10091-0>.

⁹ Per Aarvik, *op. cit.* (2019), p. 24.

actions, so there are no clear records or observable data. In this context, it becomes imperative to adopt innovative approaches to effectively prevent and detect corruption. Traditional methods have often fallen short, necessitating significant human effort¹⁰ and - as a consequence - the exploration of advanced technological solutions.

Boly and Gillanders¹¹ emphasize that reducing discretion through structured digital transformation can significantly improve institutional quality and reduce corruption. Their experimental evidence also suggests that international cooperation can facilitate the adoption of these tools, leading to more robust anti-corruption frameworks. Similarly, Bajpai and Myers¹² highlight how international cooperation has enabled governments to implement effective digital reforms and streamline administrative processes, making them more transparent and less susceptible to corruption. Building on this, coupling digital technology with robust institutional mechanisms is crucial in the fight against corruption, as it increases the cost of fraud and makes corrupt practices less attractive for actors both inside and outside government¹³.

In the digital age, both digitalization and artificial intelligence (AI) play increasingly vital roles in supporting good governance and anti-corruption efforts. Although the two concepts are often used interchangeably, they represent distinct technological processes with different implications. It is therefore important to distinguish between their functions, as each contributes differently to enhancing transparency and integrity in the public sector and plays a unique role in anti-corruption strategies.

Digitalization refers to the transformation of traditional, often manual or paper-based processes into digital formats to enhance efficiency, accessibility, and transparency¹⁴. Digitalization is boosting efficiency and agility in all sectors, particularly in public one, as it enables governments to streamline workflows, reduce bureaucracy, and provide more reliable services to citizens. Digitalization plays a fundamental role by eliminating manual processes and reducing direct interactions between citizens and public official interactions that can often create

¹⁰ Silver, D., Schrittwieser, J., Simonyan, K. *et al.* „Mastering the game of Go without human knowledge”. *Nature* 550, 354–359 (2017). <https://doi.org/10.1038/nature24270>.

¹¹ Amadou Boly, Robert Gillanders (2018), „Anti-corruption policy making, discretionary power and institutional quality: An experimental analysis”, *Journal of Economic Behavior & Organization*, 152, pp. 314-327, <https://doi.org/10.1016/j.jebo.2018.05.007>.

¹² Rajni Bajpai, C. Bernard Myers (2020), *Enhancing Government Effectiveness and Transparency: The Fight Against Corruption*, Report, vol. 1, International Bank for Reconstruction and Development / The World Bank, Malaysia, p. 267, <https://documents1.worldbank.org/curated/en/235541600116631094/pdf/Enhancing-Government-Effectiveness-and-Transparency-The-Fight-Against-Corruption.pdf>, accessed on 27.05.2025.

¹³ World Bank Group Flagship Report (2025), *op. cit.*, p. 175-180,

¹⁴ Gartner (2025), Information Technology Glossary. *Digitalization*. <https://www.gartner.com/en/information-technology/glossary/digitalization>, accessed on 27.05.2025.

opportunities for corruption. By introducing electronic platforms for public procurement, tax declarations, or document management, digitalization enhances traceability, simplifies access to information, and limits the space for non-transparent practices.

For example, in Romania, the implementation of the National Electronic System for Public Procurement (SEAP/SICAP) illustrates how digitalization enhances efficiency in anti-corruption strategies. By digitalizing the entire public procurement process, SEAP has reduced face-to-face interactions between public officials and bidders, thereby lowering the risk of favoritism and bribery. The system increases transparency by making procurement data publicly accessible, allowing for greater oversight by civil society, media, and regulatory bodies. This has contributed to more competitive bidding and improved accountability in public spending. As of February 2024, Romania's Electronic Public Procurement System (SEAP) has demonstrated significant advancements in digitalizing public procurement processes. The platform is utilized by 22,427 contracting authorities and entities, along with 216,873 economic operators—comprising 212,843 from Romania, 2,833 from other EU countries, and 1,197 from non-EU countries¹⁵. This extensive adoption has enabled approximately 99% of public procurement procedures to be conducted electronically, substantially reducing opportunities for corruption and enhancing transparency in public spending. These developments underscore Romania's commitment to leveraging digitalization in public procurement to combat corruption, increase efficiency, and foster a more transparent and accountable governance framework.

4. Artificial Intelligence Role in Anti-corruption Strategies

In contrast with digitalization, *Artificial Intelligence (AI)* involves systems capable of mimicking human intelligence - such as learning, reasoning, and decision-making - based on large volumes of data (IBM, n.d.). These systems can identify patterns, predict outcomes, and automate complex tasks, making them highly valuable for detecting fraud, analyzing risks, and supporting evidence-based decision-making in public administration. Artificial intelligence (AI) brings a higher level of intervention through its ability to analyze large volumes of data, detect suspicious patterns, and flag potential corrupt behaviors in real time. For example, in the fight against corruption, AI can be used to identify single-bid tenders, detect conflicts of interest, and assess the likelihood of fraud based on historical contract data. Thus, AI serves not only as a preventive tool but also as a proactive mechanism for early detection and automated intervention in high-risk scenarios.

¹⁵ United Nations Office on Drugs and Crime (UNODC). (2024). *Romania: Contribution to the UNCAC Working Group on Prevention (CU2024-132)*. Retrieved from https://track.unodc.org/uploads/documents/UNCAC/WorkingGroups/workinggroup4/2024-September-3-6/Contributions/CU2024-132/Romania_EN.pdf, accessed on 27.05.2025.

As artificial intelligence (AI) has evolved significantly in recent years, it is now considered an umbrella term¹⁶ that contains other concepts, such as Machine Learning (ML), Natural Language Processing (NLP) and Generative AI (GenAI).

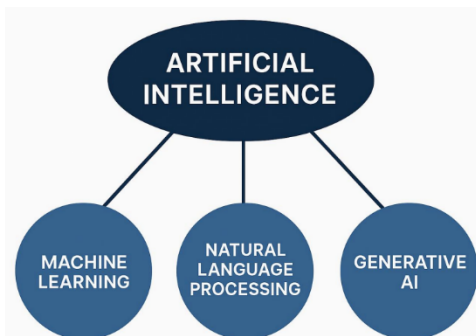


Figure 1. *Specialized Subsets of AI (ML/NLP/GenAI)*

ML models represent a subset that enables machines to learn from data and improve their performance over time without explicit programming for each task¹⁷; Natural Language Processing (NLP) enables computers to understand interpret generate and respond to human language by combining linguistics computer science and machine learning with common applications including chatbots, machine translation, sentiment analysis, speech recognition and information extraction; while GenAI models are a more specialized subset of AI, dedicated to creating new content- like text, images, or audio- by identifying and replicating patterns from existing data¹⁸. From this perspective, AI serves as the main field with the capacity to develop intelligent tools¹⁹.

As an artificial intelligence (AI)-based technology, Machine Learning (ML) enables systems to learn from data, identify patterns, and improve their performance over time without being explicitly programmed. In the context of anti-corruption, ML can process large datasets- such as public procurement records, financial transactions, or administrative logs- to detect anomalies that may signal fraudulent behavior. ML models have played a significant role in many industries to combat corruption, by detecting known patterns, irregularities and

¹⁶ Fernanda Odilla (2023), *op. cit.*, p. 38.

¹⁷ Ku. Chhaya A. Khanzode, Ravindra D. Sarode (2020), „Advantages and disadvantages of artificial intelligence and machine learning: a literature review”, *International Journal of Library and Information Science (IJLIS)*, Volume 9, Issue 1, January-April, p. 30-36, <https://doi.org/10.17605/OSF.IO/GV5T4>.

¹⁸ Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch & Patrick Zschech (2024), „Generative AI”. *Business & Information Systems Engineering* 66, 111–126. <https://doi.org/10.1007/s12599-023-00834-7>.

¹⁹ Leonardo Banh & Gero Strobel (2023), „Generative artificial intelligence”. *Electron Markets* 33, 63. <https://doi.org/10.1007/s12525-023-00680-1>.

high-risk scenarios, faster than any other known technology. For instance, ML has been effectively applied to flag irregularities like single-bid tenders or repeated contract awards, both of which are often associated with heightened corruption risk²⁰. However, the effectiveness of such technologies depends heavily on the quality of the data they rely on. If the data is biased or outdated, the technology may repeat past mistakes, fail to detect new problems, and overlook emerging patterns or missing new developments²¹, as most ML models work under imposed human criteria for specific purposes.

The PREVENT System, deployed by Romania's National Integrity Agency (ANI), exemplifies effective anti-corruption practices through the use of automation and data analysis techniques inspired by machine learning (ML). It functions as an administrative tool that monitors all ongoing public procurement procedures and contracts, relying on predefined logic and integration with multiple databases to identify potential conflicts of interest. The primary objective of the PREVENT System is to prevent such conflicts of interest by automatically detecting family ties and close connections between bidders and the management of contracting authorities²².

The application has multiple functions as follows: predictive analysis, management of investigated cases, intelligent analysis, strategic performance management and reporting, database management and decision support databases, application server and tools for technical administration, data integrator (integration with external data sources).

The National Integrity Agency (ANI) was established in 2007, in order to identify, prevent and combat integrity incidents and implemented the PREVENT system, following the adoption of Law no. 184/2016 on the establishment of a mechanism to prevent conflicts of interest in the procedure for the award of public procurement contracts.

Since 2017, through the PREVENT system, the National Integrity Agency analyzed over 117.000 public procurement procedures and issued 197 integrity warnings for potential conflicts of interest in public procurement procedures, amounting to approx. 1.9 billion Euros²³.

The added value of PREVENT System as a comprehensive monitoring

²⁰ Erica Bosio, Simeon Djankov, Edward Glaeser & Andrei Shleifer (2022), „Public Procurement in Law and Practice”, „American Economic Review”, vol. 112, no. 4, April, pp. 1091–1117.

²¹ Isabelle Adam, Mihály Fazekas (2021), *Are emerging technologies helping win the fight against corruption? A review of the state of evidence*, „Information Economics and Policy”, Volume 57, December, 100950, <https://doi.org/10.1016/j.infoecopol.2021.100950>.

²² ANI – National Integrity Agency. (2024). *PREVENT System overview*. Retrieved from: <https://integritate.eu>; United Nations Office on Drugs and Crime (UNODC). (2024). *Romania: Contribution to the UNCAC Working Group on Prevention (CU2024-132)*. Retrieved from https://track.unodc.org/uploads/documents/UNCAC/WorkingGroups/workinggroup4/2024-September-3-6/Contributions/CU2024-132/Romania_EN.pdf, accessed on 27.05.2025.

²³ ANI, 2024, *op. cit.*

in public procurement lies in its ability to monitor all public procurement procedures, in contrast to traditional oversight mechanisms that rely on sample-based verification. Sample-based systems typically audit only a portion of contracts — often randomly — leaving significant room for corrupt or high-risk transactions to go unnoticed. By comparison, PREVENT’s comprehensive coverage ensures that every procurement procedure is systematically reviewed, dramatically reducing oversight blind spots and the likelihood of undetected irregularities.

Moreover, the PREVENT System functions as an early warning mechanism, aiming to identify and address potential conflicts of interest before contracts are signed. This proactive approach marks a clear departure from the reactive nature of sample-based audits, which often intervene after integrity breaches have already occurred and caused damage.

Another important advantage of the PREVENT System is its systematic and consistent application of integrity rules. By analyzing 100% of procedures based on predefined legal criteria, the system promotes fairness and uniformity, thereby reducing the risk of selective enforcement, political favoritism, or human error in oversight.

Finally, this high level of automation and coverage contributes not only to improved legal compliance but also to greater public trust in procurement processes. In environments where corruption is a persistent challenge, the knowledge that all contracts are subject to the same integrity checks can enhance confidence among citizens, institutions, and external partners such as the European Union²⁴. While sample-based systems may appear more cost-efficient in the short term, their limited reach can render them less effective in preventing corruption-related losses.

The PREVENT System shows how key principles from machine learning— such as automation, data integration, and pattern detection—can be effectively applied to fight corruption. It demonstrates that investing in comprehensive, real-time monitoring systems can deliver higher long-term value by preventing the misuse of public funds and supporting broader goals like transparency, accountability, and institutional resilience.

In conclusion, digitalization creates the transparent and efficient framework in which corruption can be limited, while AI provides advanced analytical tools to actively identify and combat corrupt practices. Together, these technologies have the potential to fundamentally transform how public administrations prevent and respond to

Understanding the distinction between these two concepts is important for designing effective anti-corruption strategies. While digitalization lays the groundwork for transparency and accountability, AI adds an advanced layer of

²⁴ Open Government Partnership. (2023). *Romania’s commitments to open contracting*. Retrieved from <https://www.opengovpartnership.org/members/romania/>, accessed on 27.05.2025.

analytics and predictive capacity, enabling proactive governance and early intervention in corruption-prone processes.

5. Ethical Considerations in the Use of AI for Combatting Corruption

It is clear that the use of AI brings significant benefits in the fight against corruption but there are also important ethical concerns surrounding the use of AI and new technologies in anti-corruption processes that cannot be overlooked.

One main challenge is the *transparent use* of AI due to its "black box" nature which makes it difficult to explain AI-generated decisions to the public or decision makers. The key issue is developing complex AI or machine learning tools that can still be explained in a clear and simple way to regulators or auditors highlighting the broader challenge of ensuring transparency in AI deployment.

The opacity of AI decision-making processes often makes it difficult to understand how conclusions are reached, especially in high-stakes contexts like fraud detection or public procurement. This lack of explainability can hinder institutional trust and obstruct oversight. Misuse of AI and overdependence are also relevant concerns as these technologies become more embedded in governance systems. Ensuring ethical AI use requires governance frameworks that support oversight responsibility and traceability in algorithmic systems.

In my opinion a highly effective strategy to increase the transparent use of AI involves adopting a multi-level explanation framework tailored to the needs and technical backgrounds of various stakeholders, that provided a practical and effective approach for communicating AI controls to both regulators and internal decision-makers. This strategy consists of offering three levels of explanation: (1) first a simple non-technical overview of the AI model and its results in no more than two pages designed for general understanding; (2) second a more detailed and technical explanation of around nine pages that delves into how the models function for audiences with greater technical literacy; and (3) third if necessary the full algorithmic code intended for deep audits or expert scrutiny.

Regulators and decision-makers can accept the first or second format which not only satisfies regulatory requirements but also builds confidence in the technology's use. This method can also secure internal support by aligning the presentation of AI systems with a risk-based evaluation approach and offering transparency at multiple levels. This strategy effectively addresses concerns about the opaque nature of AI and supports broader and more confident adoption. It meets regulatory demands for clarity and accountability and facilitates smoother internal adoption processes offering a practical solution to the ethical and operational challenges associated with AI in regulated industries. Moreover, it ensures transparency in the use of training data and its sources and clarifies how

the algorithm is designed and implemented²⁵.

Responsible and fair use of AI also emerge as central ethical issues for the application of AI in anti-corruption efforts. A biased AI system can produce unfair outcomes such as favoring certain groups over others or reinforcing existing inequalities-outcomes that contradict the principles of responsibility and fairness. In practice, a biased AI system might treat some groups more harshly or more favorably than others in fraud detection. In anti-corruption work this could mean wrongly focusing on certain people or areas because of incorrect assumptions which can lead to unfair treatment and make people lose trust in the system. In the academic literature are explored in detail those ethical challenges, discussing the potential for AI systems to exacerbate existing biases²⁶ and the need for robust regulatory frameworks to address these risks²⁷. The complexity of AI decision-making processes, combined with the potential for biased outputs, underscores the need for ethical guidelines and regulatory oversight to ensure that AI is used responsibly in anti-corruption efforts.

Also, bias in AI systems is directly linked to the concept of responsible and fair use of AI because eliminating or minimizing bias is essential to ensuring that AI operates fairly and ethically. This aligns with ongoing debates in AI ethics emphasizing the risk of unjust targeting and the erosion of trust in systems designed to prevent corruption²⁸. Therefore, addressing bias is a core component of using AI responsibly especially in sensitive areas like anti-corruption where decisions must be impartial transparent and just. Responsible and fair use of AI means proactively identifying mitigating and monitoring bias throughout the AI lifecycle to ensure that systems support ethical goals rather than undermine them.

A contrasting perspective suggests that excessive focus on bias may unnecessarily constrain the potential of AI. This view reflects a broader debate within the AI field where ethical considerations must be weighed against opportunities for innovation and impact²⁹.

Additional ethical concerns include the phenomenon of “*AI hallucinations*” where generative models produce false or misleading outputs. “*AI hallucina-*

²⁵ European Data Protection Supervisor, EDPS MANDATE 2020 – 2024, Retrieved from chrome extension://efaidnbmnnnibpcajpcgclefindmkaj/https://www.edps.europa.eu/system/files/2025-03/25-03-06-edps-mandate-review_en.pdf, accessed on 27.05.2025.

²⁶ Allan Dafoe (2018). *AI Governance: A Research Agenda* (1442: 1443). Governance of AI Program, Future of Humanity Institute, University of Oxford, p. 5, <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf>, accessed on 27.05.2025.

²⁷ Max Tegmark (2018), *Life 3.0. Being Human in the Age of Artificial Intelligence*, Penguin Random House, p. 75.

²⁸ Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, Luciano Floridi (2016). „The ethics of algorithms: Mapping the debate”. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>.

²⁹ Erik Brynjolfsson, Danielle Li, Lindsey R. Raymond (2023), *Generative AI at Work*, NBER Working Paper No. 31161, p. 40, <http://www.nber.org/papers/w31161>, accessed on 27.05.2025.

nations" is a term used to describe instances where AI generate incorrect or unreasonable outputs or produce information that is not grounded in reality. In anti-corruption contexts this could lead to serious consequences such as the misidentification of individuals as high-risk actors resulting in reputational damage or unwarranted investigations. To address this concern, it is crucial to implement rigorous validation processes and continuous monitoring of AI systems to detect and correct such hallucinations before they can cause significant damage.

To address the ethical risks mentioned above regulatory frameworks are essential for guiding the responsible and ethical use of AI in anti-corruption strategies. Clear standards are critical, and policies focused on bias mitigation, transparency, accountability, and ethical use are necessary to ensure fairness and trust. Strong structures for ethical compliance remain crucial as they provide the foundation for ensuring that AI systems are designed, implemented, and monitored in ways that align with legal standards, social values, and human rights. These structures help organizations embed ethical principles into every stage of the AI lifecycle from data collection and model development to deployment and oversight. Without such frameworks there is a greater risk of misuse, lack of accountability and erosion of public trust especially in sensitive areas like anti-corruption where fairness and transparency are essential³⁰.

6. Where Do We Go Next? Recommendations for Further Research

Predicting the future of technology remains a complex task, especially in rapidly evolving fields like AI. While new frameworks will emerge, and some trends may fade the core principles that have guided responsible innovation, transparency collaboration adaptability and ethical design will remain essential. As efforts to fight corruption increasingly integrate AI tools the focus must stay on building systems that are not only technically advanced but also fair understandable and responsive to real-world needs. Future research and development should aim to strengthen these foundations ensuring that technology continues to serve people with simplicity stability and long-term impact.

To close this article the following section offers recommendations for further research highlighting key areas where deeper investigation could enhance the understanding and responsible application of AI in anti-corruption efforts.

For future research one area worth further exploration is the impact of AI in different contexts which would provide a more comprehensive understanding of its potential and limitations across diverse settings. This insight could support the development of more tailored and effective anti-corruption strategies. Special attention could be given to the use of AI in whistleblower protection and open

³⁰ Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, Luciano Floridi (2016), *op. cit.*

government processes where their thoughtful application could enhance transparency strengthen trust and reinforce integrity in public institutions.

Another promising area for research is the exploration of how AI can be integrated with other emerging technologies, such as blockchain and big data analytics, to enhance anti-corruption efforts. This research would provide valuable insights into the synergies between different technologies and how they can be leveraged to combat corruption more effectively.

Finally, future research should also consider the role of AI in fostering international cooperation in anti-corruption efforts. Given the global nature of corruption, AI could play a significant role in facilitating cross-border collaboration and information sharing among different countries and international organizations. Studies could explore how AI technologies can be used to enhance international cooperation, improve mutual legal assistance, and streamline cross-border investigations. Future research could also evaluate how cross-border data-sharing initiatives could improve the accuracy and reliability of predictive AI models in detecting corruption across jurisdictions.

Bibliography

1. Aarvik, Per (2019). *Artificial Intelligence – a promising anti-corruption tool in development settings?*, Bergen: CMI - CHR Michelsen Institute, (U4 Report 2019:1), <https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings>, accessed on 27.05.2025.
2. Adam, Isabelle & Mihály Fazekas (2021), *Are emerging technologies helping win the fight against corruption? A review of the state of evidence*, „Information Economics and Policy”, Volume 57, December, 100950, <https://doi.org/10.1016/j.infoecopol.2021.100950>.
3. ANI – National Integrity Agency. (2024). *PREVENT System overview*. Retrieved from: <https://integritate.eu>.
4. Bajpai, Rajni & C. Bernard Myers (2020), *Enhancing Government Effectiveness and Transparency: The Fight Against Corruption*, Report, vol. 1, International Bank for Reconstruction and Development / The World Bank, Malaysia, <https://documents1.worldbank.org/curated/en/235541600116631094/pdf/Enhancing-Government-Effectiveness-and-Transparency-The-Fight-Against-Corruption.pdf>, accessed on 27.05.2025.
5. Banh, Leonardo & Gero Strobel (2023), „Generative artificial intelligence”. *Electron Markets* 33, 63. <https://doi.org/10.1007/s12525-023-00680-1>.
6. Boly, Amadou & Robert Gillanders (2018), „Anti-corruption policy making, discretionary power and institutional quality: An experimental analysis”, *Journal of Economic Behavior & Organization*, 152, pp. 314-327, <https://doi.org/10.1016/j.jebo.2018.05.007>.
7. Bosio, Erica, Simeon Djankov, Edward Glaeser & Andrei Shleifer (2022), „Public Procurement in Law and Practice”, „American Economic Review”, vol. 112, no. 4, April, pp. 1091–1117.
8. Brynjolfsson, Erik, Danielle Li & Lindsey R. Raymond (2023), *Generative AI at Work*, NBER Working Paper No. 31161, <http://www.nber.org/papers/w31161>.

- 161, accessed on 27.05.2025.
9. Dafoe, Allan (2018). *AI Governance: A Research Agenda* (1442: 1443). Governance of AI Program, Future of Humanity Institute, University of Oxford, <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf>, accessed on 27.05.2025.
10. Feuerriegel, Stefan & Jochen Hartmann, Christian Janiesch & Patrick Zschech (2024), „Generative AI”. *Business & Information Systems Engineering* 66, 111–126. <https://doi.org/10.1007/s12599-023-00834-7>.
11. Gartner (2025), Information Technology Glossary. *Digitalization*. <https://www.gartner.com/en/information-technology/glossary/digitalization>, accessed on 27.05.2025.
12. Khanzode, Ku. Chhaya A. & Ravindra D. Sarode (2020), „Advantages and disadvantages of artificial intelligence and machine learning: a literature review”, *International Journal of Library and Information Science (IJLIS)*, Volume 9, Issue 1, January-April, p. 30-36, <https://doi.org/10.17605/OSF.IO/GV5T4>.
13. Köbis, Nils, Jean-François Bonnefon. & Iyad Rahwan (2021), „Bad machines corrupt good morals”. *Nature Human Behaviour* 5, 679–685. <https://doi.org/10.1038/s41562-021-01128-2>.
14. López-Iturriaga, Félix J. & Iván Pastor Sanz (2018), „Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces”. *Social Indicators Research* 140, 975–998. <https://doi.org/10.1007/s11205-017-1802-2>.
15. Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi (2016). „The ethics of algorithms: Mapping the debate”. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>.
16. Myint, U., 2000, „Corruption: Causes, consequences and cures”. *Asia-Pacific Development Journal*, 7(2), pp. 33-57.
17. Odilla, Fernanda (2023) „Bots against corruption: Exploring the benefits and limitations of AI-based anti-corruption technology”, *Crime, Law and Social Change Journal*, Volume 80(4), pp. 1-44, <https://doi.org/10.1007/s10611-023-10091-0>.
18. Open Government Partnership. (2023). *Romania's commitments to open contracting*. Retrieved from <https://www.opengovpartnership.org/members/romania/>, accessed on 27.05.2025.
19. Ousmane, Diagana & Mouhamadou Diagne (2023). *La corruption est un problème mondial pour le développement. Pour la combattre, nous avons tous un rôle à jouer*. La Tribune Afrique, June 2023, <https://www.banquemondiale.org/fr/news/opinion/2023/06/13/corruption-is-a-global-problem-for-development-to-fight-it-we-all-have-a-role-to-play>, accessed on 27.05.2025.
20. Silver, D., Schrittwieser, J., Simonyan, K. *et al.* „Mastering the game of Go without human knowledge”. *Nature* 550, 354–359 (2017). <https://doi.org/10.1038/nature24270>.
21. Tegmark, Max (2018), *Life 3.0. Being Human in the Age of Artificial Intelligence*, Penguin Random House.
22. Transparency International. (2024). *Corruption Perceptions Index 2024*. <https://www.transparency.org/en/cpi/2024/index/aze>, accessed on 27.05.2025.
23. United Nations Office on Drugs and Crime (UNODC). (2024). *Romania: Con-*

- tribution to the UNCAC Working Group on Prevention (CU2024-132)*. Retrieved from https://track.unodc.org/uploads/documents/UNCAC/WorkingGroups/workinggroup4/2024-September-3-6/Contributions/CU2024-132/Romania_EN.pdf, accessed on 27.05.2025.
24. World Bank Group Flagship Report (2025), *Global Economic Prospects*, International Bank for Reconstruction and Development/The World Bank, Washington, DC, <https://openknowledge.worldbank.org/server/api/core/bitstreams/f983c12d-d43c-4e41-997e-252ec6b87dbd/content>, accessed on 27.05. 2025.

CYBERSECURITY AND DIGITAL INFRASTRUCTURE RESILIENCE

The European Cybersecurity Framework: Challenges, Legal Aspects and Regulations

PhD. candidate **Leonidas SOTIROPOULOS**¹

Abstract

This article analyzes the European Union's cybersecurity evolution, tackling the dual imperatives of fostering technological advancement and ensuring systemic resilience amid rising cyber risks. Centered on the question "How do EU legislative and institutional adaptations safeguard digital sovereignty, critical infrastructure, and cross-border coordination?", it employs dogmatic legal analysis to evaluate supranational laws (NIS 1/2, Cyber Resilience Act, DORA), institutional upgrades (ENISA, CERT-EU), and policy innovations. The paper's objectives are: Transitioning from fragmented policies to a unified "cyber shield"; balancing regulatory rigor with adaptive enforcement; identifying gaps in mitigating human-centric threats and cloud vulnerabilities. The article begins with cyberspace's conceptual foundations and EU regulatory milestones. Subsequent parts dissect ENISA's capacity-building initiatives, NIS 2's expanded sectoral coverage, and the Cyber Solidarity Act's crisis-response mechanisms. Case studies on ransomware and election interference highlight systemic vulnerabilities. The conclusion underscores integration (unified threat detection), innovation (AI defenses, quantum encryption), and inclusivity (global partnerships) as pillars for maintaining Europe's leadership in ethical digital governance. By prioritizing workforce development, AI-driven solutions, and transnational collaboration, the EU seeks to establish a global standard for a resilient cybersecurity framework

Keywords: cyber-security, NIS 2, ENISA, cyber threats.

JEL Classification: K24, K33

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/04>

Please cite this article as:

Sotiropoulos, Leonidas, „The European Cybersecurity Framework: Challenges, Legal Aspects and Regulations”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 57-71.

¹ Leonidas Sotiropoulos - European University of Cyprus, LL.M in Shipping Law (Cardiff University), LLB in Law (Aristotle University of Thessaloniki in Greece), ORCID: 0009-0004-1596-7887, leon.sotiropoulos@gmail.com.

1. Introduction

The rapid evolution of cyberspace—a concept rooted in mid-20th-century cybernetics and popularized by William Gibson's vision of a "consensual hallucination"² — has transformed from a theoretical abstraction into a cornerstone of modern societal infrastructure.³ As defined in the EU's Cybersecurity Act (Regulation 2019/881)⁴, this domain encompasses the interconnected networks enabling global communication, commerce, and governance, yet its borderless nature exposes systemic vulnerabilities to increasingly sophisticated threats. A cyber threat was identified as any potential circumstance, event, or action that could destroy, disrupt, or otherwise adversely affect network and information systems, their users, or other individuals. These definitions are also referenced in the NIS 2 Directive, which we will discuss in detail below. Against this backdrop, the European Union faces a critical juncture: balancing technological innovation with the imperative to protect digital sovereignty, secure critical infrastructure, and harmonize cross-border defenses.

Basically, the term cyberspace was defined many years ago as the amorphous, hypothetical "virtual" world created by connections between computers, internet-enabled devices, servers, routers, and other elements of the internet infrastructure. However, unlike the internet itself, cyberspace is the realm generated by these connections. It exists, according to some, beyond and without any specific nation-state. The word cyberspace is a combination of the prefix "cyber-" and the word "space."⁵ The word "space" refers to a place or area and, in the context of cyberspace, denotes the virtual world within computer networks. This world is accessible through computers and other electronic devices and can be used for various purposes, such as communication, entertainment, commerce, and education. It is, therefore, an ever evolving and expanding domain, likely to play an increasingly significant role in our lives in the coming years.

The term cyberspace has existed for decades, dating back to the 1940s

² See analytically at Sabine Heuser (2003). "William Gibson's Construction of Cyberspace". In *Virtual Geographies*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004334373_005. Also, at Arulmurugan, S., and Jinnah, A.M.A., (2021). „The Cyberpunk Elements in William Gibson's *Neuromancer*". *Journal of Language and Linguistic Studies*, 17(3), 2558-2565.

³ In the 1980s, novelist William Gibson combined the prefix with space in his novel "Neuromancer," creating the term as we know it today. Gibson defined cyberspace as "...a consensual hallucination experienced daily by billions of operators, in every nation... A graphical representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data."

⁴ Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No. 526/2013 (Cybersecurity Act).

⁵ The prefix "cyber-" originates from the Greek word "kybernetes," meaning governor or pilot, and implies foresight and control.

when MIT mathematician Norbert Wiener coined the term "cybernetics." Wiener borrowed the ancient Greek adjective "kybernetikos," meaning governing, piloting, or skilled at the helm, to describe the futuristic idea that one day we would have self-regulating computer systems operating solely through feedback. In his book "Cybernetics or Control and Communication in the Animal and the Machine"⁶, the term was used to refer to the control of complex systems in the animal world and mechanical networks, particularly self-regulating control systems. Since then, cyberspace has been used by politicians, scholars, artists, and spies. It has been associated with concepts ranging from warfare to everyday online shopping, signifying both opportunities and threats.⁷

This article examines how the EU's legislative and institutional adaptations — spanning regulatory modernization, capacity-building mechanisms, and crisis-response frameworks — collectively address these challenges while navigating the tension between regulatory rigor and operational flexibility. Methodologically, the analysis employs a dogmatic legal framework to dissect supranational instruments such as the NIS 2 Directive, the Digital Operational Resilience Act (DORA), and the Cyber Resilience Act, while contextualizing their implementation through technical assessments of emerging threats (e.g., AI-enhanced ransomware) and geopolitical evaluations of cybersecurity as a tool for global leadership. Diverging from prior studies that compartmentalize technical, legal, or policy dimensions, this work adopts an integrative tripartite lens. First, it scrutinizes the technical realities of cloud vulnerabilities and hybrid warfare tactics, exemplified by the 2023 surge in state-sponsored election interference. Second, it evaluates the legal ramifications of expanded sectoral coverage under NIS 2, which now mandates cybersecurity protocols for entities ranging from energy grids to pharmaceutical manufacturers. Third, it analyzes the geopolitical implications of the EU's Cyber Solidarity Act, positioning cybersecurity as both a defensive mechanism and a vehicle for asserting normative influence in global digital governance.

The paper's originality lies in its systematic synthesis of three often-disconnected domains: legislative evolution, institutional interoperability, and threat actor innovation. By tracing the EU's transition from fragmented national policies to the envisioned "cyber shield" paradigm, it reveals how regulatory instruments like DORA's 24-hour incident reporting requirements coexist with adaptive governance structures such as AI-driven Security Operations Centres (SOCs). Concurrently, the article identifies persistent gaps, particularly in mitigating human-centric risks — evidenced by social engineering attacks compromising 68% of EU critical infrastructure breaches in 2024 — and supply chain

⁶ Norbert Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, MIT Press, 1948.

⁷ James Shires and Max Smeets (2017), *The Word Cyber Now Means Everything — and Nothing At All*, <https://slate.com/technology/2017/12/the-word-cyber-has-lost-all-meaning.html#:~:text=In%20the%201980s%2C%20novelist%20William,and%20laymen%2C%20artists%20and%20spies>

vulnerabilities exacerbated by third-party IoT device integration.⁸

Ultimately, this article contends that the EU's cybersecurity strategy hinges on three pillars: integration of threat intelligence across member states, innovation in quantum encryption and AI-driven anomaly detection, and inclusivity through partnerships with non-EU CERTs and Global South nations. By prioritizing workforce development programs to address the cybersecurity talent shortfall by 2026 and institutionalizing ethical AI governance frameworks, the EU aims to establish a global benchmark for resilient digital ecosystems.⁹ These efforts not only safeguard Europe's critical infrastructure but also position the bloc as a normative architect in the contested arena of global cyber diplomacy.

2. The European Cyber-Legislative Evolution

2.1. The European Cybersecurity Framework: Historical Context

The history of cybersecurity in the European Union (EU) is characterised by a growing awareness of the importance of digital security, reflected in evolving regulations and continuous efforts towards collaboration and reciprocity among member states. Among the early (and highly significant) initiatives of the EU was the timely recognition of the need to target an adequate level of cybersecurity, with the first step being the establishment of the European Network and Information Security Agency (ENISA) in 2004. ENISA (renamed the European Union Agency for Cybersecurity in 2019) essentially took on the task of improving network and information security across the EU, providing guidance and recommendations for the development and oversight of infrastructures and systems, depending on their criticality to both the European economy and the protection of fundamental rights of Europeans.

The European Commission early on introduced regulatory interventions, largely formulated through supranational legal instruments, aiming to achieve a unified approach and implementation across member states. The Commission published the first European cybersecurity strategy in 2013, outlining a vision for a secure and resilient cyberspace and aiming to "make the EU's digital environment the safest in the world." This was followed by Directive 2013/40/EU on attacks against information systems. Subsequent strategies and frameworks, such as the EU's 2020 cybersecurity strategy, aimed to further strengthen the EU's stance in cyberspace.

⁸ <https://www.enisa.europa.eu/sites/default/files/2024-11/2024%20Report%20on%20the%20State%20of%20the%20Cybersecurity%20in%20the%20Union.pdf>, accessed on 10.05.2025.

⁹ Lee A. Bygrave (2025), „The emergence of EU cybersecurity law: A tale of lemons, angst, turf, surf and grey boxes“, *Computer Law & Security Review*, Volume 56, 106071, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2024.106071>.

In 2016, the Network and Information Security (NIS) Directive¹⁰ was issued, marking the first EU-wide legislation to establish principles and aim to ensure a common level of cybersecurity across critical sectors such as energy, transport, finance, and healthcare. Additionally, Regulation (EU) 2019/881¹¹ introduced cybersecurity certification schemes. Member states were required to transpose this Directive into their national legislation. On 15 September 2021, Ursula von der Leyen announced in her State of the Union address that Europe, where cyber defence tools are being developed,¹² needs a European Cyber Defence Policy,¹³ including legislation on common standards based on a new European Cyber Resilience Act,¹⁴ which addresses horizontal cybersecurity requirements for products with digital elements. "If everything is connected, everything can be hacked. Given that resources are scarce, we must pool our forces. And we must not only be satisfied with addressing the threat in cyberspace but also strive to become leaders in cybersecurity. It should be here in Europe where cyber defence tools are developed, which is why we need a European Cyber Defence Policy, including new legislation on common standards based on a new European Cyber Resilience Act."

Greece for instance, has already taken a series of significant initiatives in response to international and EU requirements, shaping a secure environment for new technologies and increasing the trust of citizens and businesses in digital applications and services for the benefit of the economy and society. The issuance of Ministerial Decision 1027/2019 (Government Gazette B' 3739) on the framework of obligations for Operators of Essential Services (OES) and Digital Service Providers (DSP) marked a critical step forward in creating the network of relationships between self-regulation and oversight, essential for ensuring an adequate level of cybersecurity in critical infrastructures and services. As highlighted in the latest edition of the National Cybersecurity Strategy, "*continuous adaptation, prevention, and timely response to the challenges of an ever-changing environment form the strongest foundation for the effective shaping of a comprehensive strategy to address cyberattacks [...] and make it necessary to immediately*

¹⁰ Directive 2013/40/EU of the European Parliament and of the Council of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA.

¹¹ Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No. 526/2013 (Cybersecurity Act).

¹² Kasper, A., Mölder, H. (2020). „The EU's Common Security and Defence Policy in Facing New Security Challenges and Its Impact on Cyber Defence". In: Ramiro Troitiño, D., Kerikmäe, T., de la Guardia, R., Pérez Sánchez, G. (eds.) *The EU in the 21st Century*. Springer, Cham. https://doi.org/10.1007/978-3-030-38399-2_15.

¹³ Bendiek, Annegret (2012): *European cyber security policy*, SWP Research Paper No. RP 13/2012, Stiftung Wissenschaft und Politik (SWP), Berlin, <https://www.swp-berlin.org/publikation/european-cyber-security-policy>, accessed on 10.05.2025.

¹⁴ <https://www.european-cyber-resilience-act.com/>, accessed on 10.05.2025.

evaluate and provide feedback on the strategic planning for the country's cybersecurity."¹⁵ Moreover, the issuance of the NIS 2 Directive, which will replace the NIS Directive (hereafter referred to as NIS 1 for distinction), and the imminent publication of the Greek implementation framework will evidently require a re-assessment of the NIS 1 criteria as we move towards a much broader scope of application of its obligations, extending beyond the critical infrastructure operators covered by NIS 1.

2.2. ENISA: Institutional Backbone of EU Cybersecurity

The European Union Agency for Cybersecurity (ENISA)¹⁶ actively contributes by providing expertise, guidelines, and recommendations to member states to strengthen their capabilities in the field of cybersecurity. The main ways to achieve its objectives include:

- **Collaboration and information sharing.**
- **Strengthening communities:** As ENISA emphasises, cybersecurity is a shared responsibility.¹⁷ Europe aims to create a cross-sectoral framework for collaboration without exclusions. For this reason, ENISA has developed the **EU Cybersecurity Institutional Map** to identify and promote key stakeholders.
- **Cybersecurity policy:** According to ENISA, cybersecurity policy should not be limited to a specialised community of technical experts in cyberspace but should encompass a wide range of policy areas and initiatives.
- **Capacity building:** The demand for knowledge and skills in cybersecurity exceeds supply. The EU must invest in building capacities and talents in cybersecurity at all levels, from non-specialists to highly skilled professionals.
- **Trusted solutions:** In the process of evaluating the security of digital solutions and ensuring their reliability, a common approach must be adopted to balance the needs of society, the market, the economy, and cybersecurity. The establishment of a neutral entity acting transparently will increase customer trust in digital solutions and the broader digital environment.
- **Proactiveness:** Through a structured process enabling dialogue among stakeholders, decision-makers and policymakers will be able, on the one hand, to define strategies for timely mitigation, improving the EU's resilience to cybersecurity threats, and, on the other hand, to find solutions to emerging challenges.
- **Knowledge:** The driving force of cybersecurity is information and knowledge, necessitating a continuous process of collecting, organising, summarising, analysing, disseminating, and preserving information and knowledge about cybersecurity.

¹⁵ See https://mindigital.gr/wp-content/uploads/2022/11/E%CE%9D-NATIONAL-CYBER-SECURITY-STRATEGY-2020_2025.pdf, accessed on 10.05.2025.

¹⁶ https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/european-union-agency-cybersecurity-enisa_en, accessed on 10.05.2025.

¹⁷ Definition of Cybersecurity Gaps and overlaps in standardization, v1.0 | December 2015.

The reality is that the complexity of risks associated with information and communication technologies (ICT) is constantly increasing, and the frequency of incidents in cyberspace is rising alongside their potentially significant negative impacts.¹⁸ Moreover, due to the interconnectedness of critical sectors of the economy and state structures, ICT-related incidents may cause potential systemic effects. For this reason, managing the so-called ICT risk is of fundamental importance for an organisation to achieve its strategic, corporate, and operational objectives and safeguard its reputation. The general objectives of cybersecurity (should) include the following: availability, integrity (which may include authenticity and non-repudiation of data), and confidentiality.

Cybercrime includes all criminal offences committed using computers and communication networks. When the internet is used, it is referred to as cybercrime.¹⁹ Cybercrime via the internet primarily targets data access, illegal data trafficking, financial exploitation, or extortion of the data controller and can take the form of a generalised cyberattack aimed at disabling or disrupting networks to weaken a market or demand ransom (ransomware).²⁰ However, cybersecurity does not primarily refer to cybercrime, which falls under the jurisdiction of the Cybercrime Prosecution, but rather to its broader form, which concerns critical state infrastructures, electronic communications, physical security of infrastructures, economic activity in critical market sectors, and, ultimately, the State itself.

2.3. Legislative Framework and Modernization - the Evolution of EU Cybersecurity Framework

The NIS Directive (NIS 1),²¹ adopted in 2016, laid the foundation for a unified approach to cybersecurity across the European Union. It mandated Member States to develop national strategies for safeguarding critical network and information systems, established a Cooperation Group to facilitate cross-border collaboration, and created a network of Computer Security Incident Response

¹⁸ See analytically, Ramjee Prasad, Vandana Rohokale (2020). *Cyber security: the lifeline of information and communication technology*. Cham, Switzerland: Springer International Publishing, p. 74.

¹⁹ Dupont, B., Fortin, F., & Leukfeldt, R. (2024). „Broadening our understanding of cybercrime and its evolution.” *Journal of Crime and Justice*, 47(4), 435–439. <https://doi.org/10.1080/0735648X.2024.2323872>.

²⁰ See Andrew Jenkinson, (2022). *Ransomware and Cybercrime* (1st ed.). CRC Press. <https://doi.org/10.1201/9781003278214>. Also, Sarah Gordon & Richard Ford (2006), „On the definition and classification of cybercrime”. *Journal in Computer Virology* 2, 13–20. <https://doi.org/10.1007/s11416-006-0015-z>.

²¹ Charlotte Ducuing, „Understanding the rule of prevalence in the NIS directive: C-ITS as a case study”, *Computer Law & Security Review*, Volume 40, 2021, 105514, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2020.105514>.

Teams (CSIRTs) to address cyber threats.²² Crucially, it imposed security obligations on operators of essential services (e.g., energy, transport) and digital service providers, requiring them to adopt robust safeguards and report incidents promptly. While groundbreaking, NIS 1's limited sector coverage and fragmented enforcement highlighted the need for modernization in an era of escalating cyber risks.

In response, the NIS 2 Directive,²³ effective since January 2023, expands and strengthens this framework.²⁴ It broadens the scope to include 13 critical sectors — up from 11 under NIS 1 — such as public administration, space, and wastewater management. Digital service providers, including online marketplaces and cloud platforms, now face stricter compliance requirements. NIS 2 introduces enhanced risk management protocols, mandating organizations to conduct comprehensive cybersecurity assessments, implement mitigation measures like supply chain audits, and report incidents to National Cybersecurity Authorities (NCAs) within 24 hours. New obligations include cybersecurity training for employees, public awareness campaigns, and formal incident response plans. All Member States transposed NIS 2 into national law, replacing NIS 1 entirely.

The Digital Operational Resilience Act (DORA),²⁵ targeting the financial sector, mandates stringent ICT risk management for banks, insurers, and fintech firms. It requires regular stress testing, third-party vendor oversight, and real-time incident reporting to ensure operational continuity during cyberattacks.²⁶ Parallely, the EU Cyber Resilience Act addresses vulnerabilities in connected devices (e.g., smart appliances, IoT systems).²⁷ Manufacturers must now embed cybersecurity features during product design, disclose vulnerabilities transparently, and

²² See characteristically, Pauline Meyer & Sylvain Mètille (2023), „Computer security incident response teams: are they legally regulated? The Swiss example”. *International Cybersecurity Law Review* 4, 39–60. <https://doi.org/10.1365/s43439-022-00070-x>.

²³ Niels Vandezande, „Cybersecurity in the EU: How the NIS2-directive stacks up against its predecessor”, *Computer Law & Security Review*, Volume 52, 2024, 105890, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2023.105890>.

²⁴ Paula Contreras (2023). „The Transnational Dimension of Cybersecurity: The NIS Directive and Its Jurisdictional Challenges”. In: Cyril Onwubiko, Pierangelo Rosati, Aunshul Rege, Arnau Erola, Xavier Bellekens, Hanan Hindy, Martin Gilje Jaatun, *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media*. Springer Proceedings in Complexity. Springer, Singapore. https://doi.org/10.1007/978-981-19-6414-5_18.

²⁵ Neumannová, Anita, Edward W. N. Bernroider & Christoph Elshuber (2023). „The Digital Operational Resilience Act for Financial Services: A Comparative Gap Analysis and Literature Review”. In: Maria Papadaki, Paulo Rupino da Cunha, Marinos Themistocleous, Klitos Christodoulou (eds.) *Information Systems. EMCIS 2022. Lecture Notes in Business Information Processing*, vol 464. Springer, Cham, pp. 327–341, https://doi.org/10.1007/978-3-031-30694-5_40.

²⁶ Dirk Clausmeier (2023), „Regulation of the European Parliament and the Council on digital operational resilience for the financial sector (DORA)”. *International Cybersecurity Law Review* 4, 79–90. <https://doi.org/10.1365/s43439-022-00076-5>.

²⁷ See Pier Giorgio Chiara, „The Cyber Resilience Act: the EU Commission's proposal for a horizontal regulation on cybersecurity for products with digital elements.” *International Cybersecurity Law Review* 3, 255–272 (2022). <https://doi.org/10.1365/s43439-022-00067-6>.

comply with EU-wide certification schemes before releasing goods to the market.

Regulation 2023/2841, effective since January 2024, establishes a Cybersecurity Governance Framework for EU institutions. It tasks the newly formed Cybersecurity Advisory Board (IICB) with overseeing compliance, while expanding the role of CERT-EU (the EU's central cyber incident response body) to coordinate threat intelligence sharing and crisis management. Together, these measures aim to harmonize standards, foster cross-border collaboration, and preempt emerging threats like AI-driven attacks or ransomware targeting critical infrastructure.

The EU's legislative push — spearheaded by NIS 2, DORA, and the Cyber Resilience Act — reflects a paradigm shift from reactive to proactive cybersecurity. By unifying reporting standards, broadening sectoral coverage, and institutionalizing cross-border cooperation, the bloc seeks to fortify its digital ecosystem against evolving threats. For businesses, this translates to heavier compliance burdens but also opportunities to build trust through demonstrable cyber resilience.

3. Operational Dynamics & Strategic Responses

3.1. Cybersecurity Challenges and Threat Landscape

Growing anxieties over the security of digital infrastructure and the vast quantities of data circulating across digital systems and subsystems have reached unprecedented levels, with trends indicating a persistent upward trajectory. The European Union's intensified focus on collaborative frameworks, legislative reforms, and the establishment of resilient cybersecurity architectures highlights its commitment to fortifying a secure digital environment for member states.

According to the European Union Agency for Cybersecurity (ENISA), cybersecurity challenges now threaten the foundational pillars of democratic Europe, extending beyond isolated sectors or organisations reliant on digital infrastructure. In its most recent analysis of evolving cyber threat trends, ENISA underscores a disturbing shift: cyberattacks increasingly target individuals in positions of influence, including employees in critical roles, politicians, government officials, journalists, and activists. Attackers predominantly deploy spear-phishing emails and exploit social media platforms to infiltrate systems. Notable tactics include²⁸:

- Malicious advertising campaigns, where counterfeit websites masquerade as legitimate applications, enabling attackers to hijack system boot processes and bypass security protocols.

²⁸ Peter Swire, DeBrae Kennedy-Mayo, Drew Bagley, Sven Krasser, Avani Modak, and Christoph Bausewein. (2024). "Risks to Cybersecurity from Data Localization, Organized by Techniques, Tactics and Procedures." *Journal of Cyber Policy* 9 (1): 20–51. doi: 10.1080/23738871.2024.2384724.

- Exploitation of cloud infrastructure misconfigurations, a method that not only compromises cloud-based storage, networks, and systems but also extends to cloud management consoles, granting attackers broad control over operational environments.

In 2024 Cyber threats intensified globally, marked by a 35% surge in ransomware (40% targeting healthcare/education) and 25% more phishing campaigns, with AI-generated emails boosting success rates by 50%. Critical infrastructure (energy/transportation) faced 70% of incidents, while IoT device attacks rose 60% and AI-driven threats comprised 20% of advanced attacks. Data breaches cost averaged \$4.5M (+15% YoY), with human error causing 30%. Sector-specific risks spiked: financial attacks (+50%, exploiting APIs) and healthcare breaches (+40%, targeting patient data). State-sponsored attacks grew 30%, focusing on espionage, with Europe hit by 25% of global incidents. These trends demand urgent AI-augmented defenses, infrastructure hardening, and international cooperation.²⁹

Furthermore, the escalating frequency of cloud-based vulnerabilities³⁰ underscores systemic risks, as misconfigured environments offer attackers opportunities to disrupt operations or exfiltrate sensitive data. Elections, as a cornerstone of democratic processes, face heightened risks due to attacks on public administration and essential service providers. Meanwhile, the trend toward human-centric targeting — using psychologically manipulative tactics against high-profile individuals — reflects adversaries' growing sophistication in exploiting social dynamics.

Initiatives such as the NIS 2 Directive and the Cyber Resilience Act exemplify the bloc's efforts to address these challenges through harmonised security standards, stringent compliance mandates, and enhanced cross-border collaboration.³¹ However, the rapidly evolving threat landscape demands continuous innovation in defensive technologies, investment in workforce training, and strengthened public-private partnerships to safeguard Europe's digital future.

3.2. Cyber-Strategic Initiatives and Institutional Coordination

The European Cybersecurity Competence Centre and Network (ECCC),³² established under a 2021 Regulation, represents a cornerstone of the

²⁹ <https://ciras.enisa.europa.eu/>, accessed on 10.05.2025.

³⁰ FNU Jimmy, (2024). „Cyber security Vulnerabilities and Remediation Through Cloud Security Tools”. *Journal of Artificial Intelligence General Science (JAIGS)* ISSN:3006-4023, 2(1), 129–171. <https://doi.org/10.60087/jaigs.v2i1.102>.

³¹ Philipp Eckhardt & Anastasia Kotovskaia (2023), „The EU's cybersecurity framework: the interplay between the Cyber Resilience Act and the NIS 2 Directive”. *International Cybersecurity Law Review* 4, 147–164. <https://doi.org/10.1365/s43439-023-00084-z>.

³² Sebastian Suciú & Andreea-Larisa Cirjan (2022). „The European Cybersecurity Competence Centre-One More Step towards Supranationalism”. *Perspective Politice*, vol. XV, no. 1-2, pp. 57-73, DOI: 10.25019/perspol/22.15.4.

EU's strategy to bolster cybersecurity capabilities. Headquartered in Bucharest, the ECCC collaborates with National Coordination Centres (NCCs) across member states to drive innovation, industrial policy, and resilience. By pooling resources from the EU, member states, and private industry, the Centre focuses on developing shared technological priorities for critical sectors such as public administration, energy, and SMEs. Its mission extends beyond infrastructure protection to fostering cross-border collaboration among researchers, industry leaders, and public institutions, ensuring Europe maintains its competitive edge in cybersecurity technologies while safeguarding democratic institutions and economic stability.³³

The Cyber Europe 2024 exercise, scheduled for June 2024, exemplifies the EU's proactive approach to crisis preparedness. As the 7th iteration of pan-European cybersecurity drills, it simulates a scenario involving geopolitical tensions at the EU's borders, including foreign state interference and coordinated attacks on energy infrastructure.³⁴ Secondary targets include digital service providers and public administration systems, reflecting growing concerns over hybrid threats to critical infrastructure. These exercises, building on lessons from Cyber Europe 2020, aim to refine incident response protocols and strengthen coordination among ENISA, national agencies, and private stakeholders.

The CERT-EU, now rebranded as the Cybersecurity Service for EU Institutions, has expanded its mandate to become a central hub for threat intelligence sharing and incident response. Its upgraded role includes providing advisory services to EU bodies, analyzing emerging threats, and coordinating cross-institutional responses to cyber incidents.³⁵ This evolution underscores the EU's commitment to protecting its own governance structures from increasingly sophisticated attacks, particularly those targeting sensitive political or operational data.³⁶

³³ Aljosa Pasic (2022), „Governance Mesh Approach for Cybersecurity Ecosystem”, *Information & Security: An International Journal*, vol. 53, no. 1: 105-130, <https://doi.org/10.11610/isij.5308>.

³⁴ Darlington Eze Ekechukwu, & Peter Simpa (2024). „The future of cybersecurity in renewable energy systems: A review, identifying challenges and proposing strategic solutions”. *Computer Science & IT Research Journal*, 5(6), 1265–1299. DOI: 10.51594/csitrj.v5i6.1197R. Also see Rauno Pirinen, Paresh Rathod, Emilia Gugliandolo, Kevin Fleming, Nineta Polemi, "Towards the Harmonisation of Cybersecurity Education and Training in the European Union Through Innovation Projects," *2024 IEEE Global Engineering Education Conference (EDUCON)*, Kos Island, Greece, 2024, pp. 1-9, doi: 10.1109/EDUCON60312.2024.10578867.

³⁵ Pythagoras Petratos, (2014). „Cybersecurity in Europe: Cooperation and Investment.” In: Elias G. Carayannis, David F. J. Campbell, Marios Panagiotis Efthymiopoulos, (eds.), *Cyber-Development, Cyber-Democracy and Cyber-Defense*. Springer, New York, p. 279-301, https://doi.org/10.1007/978-1-4939-1028-1_11.

³⁶ Christian Calliess, Ansgar Baumgarten (2020). „Cybersecurity in the EU the example of the financial sector: a legal perspective”. *German Law Journal*, 21(6), 1149-1179. Also see Odermatt, J. (2018). „The European Union as a Cybersecurity Actor”. In: Blockmans, S. & Koutrakos, P. (Eds.), *Research Handbook on EU Common Foreign and Security Policy*. (pp. 354-373). Cheltenham, UK: Edward Elgar Publishing. ISBN 9781785364075 doi: 10.4337/9781785364082.00026.

The EU Cyber Solidarity Act (2023), proposed in April 2023, introduces a multi-layered framework to enhance collective resilience. Central to this initiative is the European Cybersecurity Shield, a network of interconnected Security Operations Centres (SOCs) leveraging AI and data analytics to detect threats in real time.³⁷ Funded through the Digital Europe Programme, these SOCs — including three cross-border consortia involving 17 member states and Iceland — form a distributed early-warning system. Complementing this infrastructure is a **Cybersecurity Reserve**, a pool of pre-vetted private response teams available to assist member states during large-scale incidents, and a **Mutual Assistance Framework** enabling states to request or provide cross-border support during crises.

Strategic priorities under these initiatives focus on achieving **technological sovereignty** by reducing reliance on non-EU cybersecurity solutions, fostering homegrown innovation through public-private partnerships, and addressing hybrid threats through stress-testing critical sectors like finance and healthcare. By aligning the ECCC's research capabilities, CERT-EU's operational expertise, and the Cyber Solidarity Act's response mechanisms, the EU aims to create a unified "cyber shield"³⁸ capable of anticipating disruptions, mitigating attacks, and ensuring rapid recovery — a vital component of its broader digital single market and geopolitical resilience agenda.³⁹

4. Conclusion

The European Union's cybersecurity framework has transitioned from fragmented national policies to a cohesive, forward-looking strategy marked by legislative milestones like the NIS 2 Directive, DORA, and the Cyber Resilience Act, which prioritize harmonized regulations, cross-border collaboration, and technological sovereignty. Central to this evolution are institutional advancements such as the European Cybersecurity Competence Centre (ECCC) and the Cyber Solidarity Act, which pool resources and expertise to counter state-sponsored threats and critical infrastructure vulnerabilities. Despite progress, challenges persist, including AI-driven ransomware, supply chain exploits, and workforce shortages, compounded by the need for consistent implementation of directives across member states. Looking ahead, the EU aims to integrate initiatives like the Cybersecurity Shield and DORA for seamless threat response, innovate

³⁷ Pier Giorgio Chiara and Laura Bartoli, *Unveiling EU Cybersecurity Law Turf Battles: The Case of the EU Cyber Solidarity Act Proposal*. Available at SSRN: <https://ssrn.com/abstract=4719533> or <http://dx.doi.org/10.2139/ssrn.4719533>.

³⁸ Anna-Maria Osula (2022). „Building Cyber Resilience: The Defensive Shield for the EU”. In: Gertjan Boulet, Michael Reiterer, Ramon Pacheco Pardo (eds.) *Cybersecurity Policy in the EU and South Korea from Consultation to Action. New Security Challenges*. Palgrave Macmillan, Cham. pp. 179–196, https://doi.org/10.1007/978-3-031-08384-6_9.

³⁹ Izabela Oleksiewicz, Mustafa Emre Civelek, (2023). „Where are the changes in EU cybersecurity legislation leading?”. *Humanities and Social Sciences*, 30(4-part 1), 183-197.

through AI-driven defenses and quantum-resistant encryption, and strengthen global partnerships to combat transnational cybercrime. By embedding cybersecurity into digital transformation agendas and balancing regulatory rigor with adaptive governance, the bloc seeks to set a global standard for resilient, ethical digital ecosystems capable of mitigating 21st-century threats.

Bibliography

1. Arulmurugan, S. and A.M.A. Jinnah (2021). „The Cyberpunk Elements in William Gibson’s Neuromancer”. *Journal of Language and Linguistic Studies*, 17(3), 2558-2565.
2. Bendick, Annegret (2012): *European cyber security policy*, SWP Research Paper No. RP 13/2012, Stiftung Wissenschaft und Politik (SWP), Berlin, <https://www.swp-berlin.org/publikation/european-cyber-security-policy>, accessed on 10.05.2025.
3. Bygrave, Lee A. (2025), „The emergence of EU cybersecurity law: A tale of lemons, angst, turf, surf and grey boxes”, *Computer Law & Security Review*, Volume 56, 106071, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2024.106071>.
4. Calliess, Christian & Ansgar Baumgarten (2020). „Cybersecurity in the EU the example of the financial sector: a legal perspective”. *German Law Journal*, 21 (6), 1149-1179.
5. Chiara, Pier Giorgio & Laura Bartoli, *Unveiling EU Cybersecurity Law Turf Battles: The Case of the EU Cyber Solidarity Act Proposal*. Available at SSRN: <https://ssrn.com/abstract=4719533> or <http://dx.doi.org/10.2139/ssrn.4719533>.
6. Chiara, Pier Giorgio, „The Cyber Resilience Act: the EU Commission’s proposal for a horizontal regulation on cybersecurity for products with digital elements.” *International Cybersecurity Law Review* 3, 255–272 (2022). <https://doi.org/10.1365/s43439-022-00067-6>.
7. Clausmeier, Dirk (2023), „Regulation of the European Parliament and the Council on digital operational resilience for the financial sector (DORA)”. *International Cybersecurity Law Review* 4, 79–90. <https://doi.org/10.1365/s43439-022-00076-5>.
8. Contreras, Paula (2023). „The Transnational Dimension of Cybersecurity: The NIS Directive and Its Jurisdictional Challenges”. In: Onwubiko, Cyril, Pierangelo Rosati, Aunshul Rege, Arnau Erola, Xavier Bellekens, Hanan Hindy & Martin Gilje Jaatun, *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media*. Springer Proceedings in Complexity. Springer, Singapore. https://doi.org/10.1007/978-981-19-6414-5_18.
9. Directive 2013/40/EU of the European Parliament and of the Council of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA.
10. Ducuing, Charlotte, „Understanding the rule of prevalence in the NIS directive: C-ITS as a case study”, *Computer Law & Security Review*, Volume 40, 2021, 105514, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2020.105514>.

11. Dupont, Benoît, Francis Fortin & Rutger Leukfeldt (2024). „Broadening our understanding of cybercrime and its evolution.” *Journal of Crime and Justice*, 47(4), 435–439. <https://doi.org/10.1080/0735648X.2024.2323872>.
12. Eckhardt, Philipp & Anastasia Kotovskaia (2023), „The EU’s cybersecurity framework: the interplay between the Cyber Resilience Act and the NIS 2 Directive”. *International Cybersecurity Law Review* 4, 147–164. <https://doi.org/10.1365/s43439-023-00084-z>.
13. Ekechukwu, Darlington Eze & Peter Simpa (2024). „The future of cybersecurity in renewable energy systems: A review, identifying challenges and proposing strategic solutions”. *Computer Science & IT Research Journal*, 5(6), 1265–1299. DOI: 10.51594/csitrj.v5i6.1197R.
14. Gordon, Sarah & Richard Ford (2006), „On the definition and classification of cybercrime”. *Journal in Computer Virology* 2, 13–20. <https://doi.org/10.1007/s11416-006-0015-z>.
15. Heuser, Sabine (2003). "William Gibson’s Construction of Cyberspace". In *Virtual Geographies*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004334373_005.
16. Jenkinson, Andrew (2022). *Ransomware and Cybercrime* (1st ed.). CRC Press. <https://doi.org/10.1201/9781003278214>.
17. Jimmy, FNU (2024). „Cyber security Vulnerabilities and Remediation Through Cloud Security Tools”. *Journal of Artificial Intelligence General Science (JAIGS)* ISSN:3006-4023, 2(1), 129–171. <https://doi.org/10.60087/jaigs.v2i1.102>.
18. Kasper, A., H. Mölder (2020). „The EU’s Common Security and Defence Policy in Facing New Security Challenges and Its Impact on Cyber Defence”. In: Ramiro Troitiño, D., Kerikmäe, T., de la Guardia, R., Pérez Sánchez, G. (eds.) *The EU in the 21st Century*. Springer, Cham. https://doi.org/10.1007/978-3-030-38399-2_15.
19. Meyer, Pauline & Sylvain Métile (2023), „Computer security incident response teams: are they legally regulated? The Swiss example”. *International Cybersecurity Law Review* 4, 39–60. <https://doi.org/10.1365/s43439-022-00070-x>.
20. Neumannová, Anita, Edward W. N. Bernroider & Christoph Elshuber (2023). „The Digital Operational Resilience Act for Financial Services: A Comparative Gap Analysis and Literature Review”. In: Maria Papadaki, Paulo Rupino da Cunha, Marinos Themistocleous, Klitos Christodoulou (eds.) *Information Systems*. EMCIS 2022. Lecture Notes in Business Information Processing, vol 464. Springer, Cham, pp. 327–341, https://doi.org/10.1007/978-3-031-30694-5_40.
21. Odermatt, J. (2018). „The European Union as a Cybersecurity Actor”. In: Blockmans, S. & Koutrakos, P. (Eds.), *Research Handbook on EU Common Foreign and Security Policy*. (pp. 354-373). Cheltenham, UK: Edward Elgar Publishing. ISBN 9781785364075 doi: 10.4337/9781785364082.00026.
22. Oleksiewicz, Izabela & Mustafa Emre Civelek, (2023). „Where are the changes in EU cybersecurity legislation leading?”. *Humanities and Social Sciences*, 30(4-part 1), 183-197.
23. Osula, Anna-Maria (2022). „Building Cyber Resilience: The Defensive Shield for the EU”. In: Gertjan Boulet, Michael Reiterer, Ramon Pacheco Pardo (eds.) *Cybersecurity Policy in the EU and South Korea from Consultation to Action. New Security Challenges*. Palgrave Macmillan, Cham. pp. 179–196, https://doi.org/10.1007/978-1-349-70000-0_10.

- org/10.1007/978-3-031-08384-6_9.
24. Pasic, Aljosa (2022), „Governance Mesh Approach for Cybersecurity Ecosystem”, *Information & Security: An International Journal*, vol. 53, no. 1: 105-130, <https://doi.org/10.11610/isij.5308>.
25. Petratos, Pythagoras (2014). „Cybersecurity in Europe: Cooperation and Investment.” In: Carayannis, Elias G., David F. J. Campbell, Marios Panagiotis Efthymiopoulos, (eds.), *Cyber-Development, Cyber-Democracy and Cyber-Defense*. Springer, New York, p. 279-301, https://doi.org/10.1007/978-1-4939-1028-1_11.
26. Pirinen, Rauno, Paresh Rathod, Emilia Gugliandolo, Kevin Fleming & Nineta Polemi, "Towards the Harmonisation of Cybersecurity Education and Training in the European Union Through Innovation Projects," *2024 IEEE Global Engineering Education Conference (EDUCON)*, Kos Island, Greece, 2024, pp. 1-9, doi: 10.1109/EDUCON60312.2024.10578867.
27. Prasad, Ramjee & Vandana Rohokale (2020). *Cyber security: the lifeline of information and communication technology*. Cham, Switzerland: Springer International Publishing.
28. Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No. 526/2013 (Cybersecurity Act).
29. Shires, James and Max Smeets (2017), *The Word Cyber Now Means Everything — and Nothing At All*, <https://slate.com/technology/2017/12/the-word-cyber-has-lost-all-meaning.html#:~:text=In%20the%201980s%2C%20novelist%20William,and%20laymen%2C%20artists%20and%20spies>.
30. Suci, Sebastian & Andreea-Larisa Cirjan (2022). „The European Cybersecurity Competence Centre-One More Step towards Supranationalism”. *Perspective Politice*, vol. XV, no. 1-2, pp. 57-73, DOI: 10.25019/perspol/22.15.4.
31. Swire, Peter, DeBrae Kennedy-Mayo, Drew Bagley, Sven Krasser, Avani Modak and Christoph Bausewein. (2024). “Risks to Cybersecurity from Data Localization, Organized by Techniques, Tactics and Procedures.” *Journal of Cyber Policy* 9 (1): 20–51. doi: 10.1080/2373 8871.2024.2384724.
32. Vandezande, Niels, „Cybersecurity in the EU: How the NIS2-directive stacks up against its predecessor”, *Computer Law & Security Review*, Volume 52, 2024,105890, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2023.105890>.
33. Wiener, Norbert, *Cybernetics or Control and Communication in the Animal and the Machine*, MIT Press, 1948.

NIS2 Directive - Legal Preparedness of EU Health Infrastructure Against Large-Scale Cyberattacks

Ph.D. student **Antonia RENGLE**¹

Abstract

This study examines the legal preparedness of European Union health infrastructure under the NIS2 Directive (Directive (EU) 2022/2555) against large-scale cyberattacks, focusing on the health sector as critical infrastructure. Its primary objective is to assess the effectiveness of NIS2's legal mechanisms – risk management, incident reporting, and management accountability – in safeguarding health systems. The research methodology involves a detailed analysis of four recent case studies: Synnovis (2024), NailaoLocker (2024), HSE (2021/2024), and Vastaamo (2020/2024), supplemented by additional research from sources such as ENISA reports and European Commission documents. Findings highlight strengths, including rapid reporting and management accountability, alongside weaknesses such as coordination delays, legacy system vulnerabilities, and uneven transposition. The implications indicate that while NIS2 provides a robust framework, it requires operational and financial support to ensure resilience, proposing reforms like a unified crisis protocol and mandatory system upgrades. This study contributes to the legal discourse on EU cybersecurity, emphasizing the need for harmonization and adequate resources.

Keywords: NIS2 Directive, EU health infrastructure, large-scale cyberattacks, critical infrastructure, cyber vulnerabilities, management accountability.

JEL Classification: K24, K32

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/05>

Please cite this article as:

Rengle, Antonia, „NIS2 Directive - Legal Preparedness of EU Health Infrastructure Against Large-Scale Cyberattacks”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 72-91.

1. Introduction

The digital transformation of healthcare within the European Union² has

¹ Antonia Rengle - Doctoral School of Law, Babeş-Bolyai University of Cluj-Napoca, Romania, ORCID: <https://orcid.org/0009-0008-8187-686X>, rengleantonia@gmail.com.

² European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2023,” October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

entrenched its status as critical infrastructure, integrating hospitals, clinical laboratories, and pharmaceutical entities into a complex, networked ecosystem that underpins public health and societal stability.³ Over the past two decades, the adoption of electronic health records, telemedicine platforms, and interconnected medical devices has revolutionized patient care delivery, enabling faster diagnoses, remote monitoring, and streamlined operations across member states.⁴ However, this technological evolution has coincided with a marked increase in large-scale cyberattacks targeting the healthcare sector, with incidents such as ransomware and data breaches exploiting vulnerabilities in these interconnected systems to disrupt operations and compromise patient safety.⁵ The European Union Agency for Cybersecurity (ENISA) reported in its 2023 Threat Landscape that ransomware accounted for 54% of cybersecurity incidents in healthcare between July 2022 and June 2023, a statistic underscoring the sector's growing exposure to sophisticated threats.⁶ In response, the NIS2 Directive (Directive (EU) 2022/2555), enacted on November 14, 2022, and published in the Official Journal of the European Union on December 27, 2022, represents the EU's latest legislative effort to strengthen cybersecurity across essential sectors, explicitly designating healthcare as a priority due to its critical role in public welfare and the immediate human consequences of service disruptions.⁷ Mandated for transposition into national laws by October 17, 2024, NIS2 aims to enhance resilience against such threats, though as of 2025, the implementation process remains ongoing across member states, with varying degrees of progress reported by national authorities.⁸

The central research problem of this study is to determine whether NIS2's

³ European Commission, "eHealth: Digital Health and Care," accessible at https://health.ec.europa.eu/ehealth-digital-health-and-care_en, accessed March 19, 2025.

⁴ ENISA, "Checking-up on Health: Ransomware Accounts for 54% of Cybersecurity Threats," July 4, 2023, accessed at <https://www.enisa.europa.eu/news/checking-up-on-health-ransomware-accounts-for-54-of-cybersecurity-threats>, last accessed 03.03.2025.

⁵ Benyamine Abbou, Boris Kessel, Merav Ben Natan, Rinat Gabbay-Benziv, Dikla Dahan Shriki, Anna Ophir, Nimrod Goldschmid et al., (2024). "When all computers shut down: the clinical impact of a major cyber-attack on a general hospital", *Frontiers in Digital Health*, 6. Accessible at <https://doi.org/10.3389/fdgth.2024.1321485>, last revised 01.03.2025

⁶ Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) Recital 12, Article 41. Accessible at <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>, last accessed 01.02.2025.

⁷ European Commission, "Cybersecurity Policies," accessed March 19, 2025, available at: <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-policies>.

⁸ Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), Article 21, Article 23, Article 20. Accessible at <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>, last accessed 01.02.2025.

legal mechanisms—specifically risk management under Article 21, incident reporting under Article 23, and management accountability under Article 20 — are sufficient to equip EU health infrastructure to withstand large-scale cyberattacks.⁹ This question is of pressing importance given the sector’s unique vulnerabilities: outdated IT systems, often running unpatched software; limited financial and technical resources, particularly in smaller healthcare facilities; and the high stakes of patient care, where disruptions can lead to delayed treatments or compromised medical data, amplifying the human and operational impact of cyber incidents.¹⁰ Previous doctrinal analyses, such as those conducted under the NIS1 Directive (Directive (EU) 2016/1148), have typically adopted a broad approach, assessing cybersecurity compliance across multiple sectors — energy, transport, and healthcare — without a specific focus on the distinct legal and operational challenges faced by health infrastructure.¹¹ For instance, the European Commission’s 2019 assessment of NIS1 implementation highlighted general compliance issues but offered limited insight into sector-specific preparedness, leaving a gap in understanding healthcare’s unique needs.¹² This study introduces a novel perspective by concentrating exclusively on the legal preparedness of EU health infrastructure under NIS2, leveraging real, documented case studies to test its provisions and propose targeted reforms based solely on verified evidence available as of 2025.¹³ The significance of this inquiry lies in its potential to bridge theoretical legal frameworks with practical resilience, offering a focused contribution to the ongoing discourse on EU cyberspace governance — a critical area of contemporary legal scholarship amid the rising frequency and sophistication of cyber threats.¹⁴

⁹ European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2023,” October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

¹⁰ Directive (EU) 2016/1148, “Concerning Measures for a High Common Level of Security of Network and Information Systems Across the Union,” July 6, 2016, accessible at <https://eur-lex.europa.eu/eli/dir/2016/1148/oj>, last revised 02.06.2024 and Giese, Gerd and Frank Bartel, 2025. “How to secure development environments”, CSJ(3), 8:232. <https://doi.org/10.69554/jath1370>.

¹¹ European Commission, “Assessment of the EU Member States’ Rules on Cybersecurity,” July 10, 2019, accessible at <https://digital-strategy.ec.europa.eu/en/library/assessment-eu-member-states-rules-cybersecurity>, last revised 02.03.2025.

¹² Directive (EU) 2022/2555, Article 1(1). Accessible at <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>, last accessed 01.02.2025 to be named moving forward Directive (EU) 2022/2555 and Jyri Rajamäki, Dominik Jarzemski, Jiri Kucera, Ville Nyman, Ilmari Pura, Jarno Virtanen, Minna Herlevi et al., 2024. “Implications of GDPR and NIS2 for cyber threat intelligence exchange in hospitals”, *Wseas Transactions on Computers*, 23:1-11. Available at: <https://doi.org/10.37394/23205.2024.23.1>.

¹³ European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2023,” October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

¹⁴ UK National Audit Office, “Investigation: WannaCry Cyber Attack and the NHS,” April 25, 2018, accessible at <https://www.nao.org.uk/reports/investigation-wannacry-cyber-attack-and-the-nhs/>, last accessed 01.02.2025 Finnish Data Protection Ombudsman, “Decision on Vastaamo Data

The article is structured into seven sections to provide a thorough and systematic analysis. This introduction delineates the research problem, underscores the study's novelty in relation to existing literature, outlines the article's structure, details the research methodology, and previews the proposed solutions. Subsequent sections include an in-depth analysis of NIS2's legal framework, a detailed examination of three case studies to test its potential application, an assessment of its legal strengths, an identification of its weaknesses and gaps, a proposal of reforms to enhance preparedness, and a concluding section summarizing findings and their implications. The research methodology combines a case study approach with doctrinal analysis, focusing on three well-documented incidents — WannaCry (May 2017), Vastaamo (October 2020), and Synnovis (June 2024) — using only publicly available data to evaluate how NIS2's mechanisms might have applied or could apply in practice.¹⁵ These cases are selected for their relevance to large-scale cyberattacks on health infrastructure: WannaCry disrupted NHS services across the UK, Vastaamo exposed sensitive patient data in Finland with cross-border ramifications, and Synnovis impacted pathology services in London, all providing concrete examples of the threats NIS2 aims to address.¹⁶ The analysis is supplemented by a detailed examination of NIS2's legislative text, as published on EUR-Lex, and secondary sources, including ENISA's 2023 Threat Landscape report, the UK National Audit Office's WannaCry investigation, and official statements from Finnish and NHS authorities.¹⁷ Proposed solutions, informed by these cases, include establishing a unified crisis protocol to improve coordination and mandating system upgrades to address legacy vulnerabilities, aiming to enhance both the legal and operational resilience of EU health infrastructure.¹⁸ This study seeks to provide actionable insights for EU policymakers, ensuring that health systems are adequately protected against the growing threat of large-scale cyberattacks, thereby advancing the legal framework for cybersecurity within the Union as NIS2 takes effect.

Breach," October 29, 2020, accessible at <https://tietosuojafi/en/-/decision-on-vastaamo-data-breach>, last revised 01.02.2025 ,NHS England, "NHS Statement on Synnovis Cyber Incident," June 4, 2024, accessible at <https://www.england.nhs.uk/news/>, last revised 02.03.2025

¹⁵ European Commission, "Cybersecurity Strategy for the Digital Decade," December 16, 2020, accessible at <https://digital-strategy.ec.europa.eu/en/library/eus-cybersecurity-strategy-digital-decade-0>, last revised 02.03.2025.

¹⁶ Directive (EU) 2022/2555, Article 2. Accessible at <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>, last accessed 01.02.2025 and Robert Mikac, 2023. "Protection of the EU's Critical Infrastructures: Results and Challenges", *Applied Cybersecurity & Internet Governance* (1), 2:1-5. <https://doi.org/10.60097/acig/162868>.

¹⁷ Directive (EU) 2022/2555, Article 41. Accessible at <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>, last accessed 01.02.2025 and Robert Mikac, *op. cit.*, pp. 1-5.

¹⁸ Directive (EU) 2022/2555, Annex I. Accessible at <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>, last accessed 01.02.2025.

2. Legal Framework of NIS2

The NIS2 Directive represents a significant evolution from the 2016 NIS Directive (Directive (EU) 2016/1148), establishing a robust and comprehensive legal framework aimed at addressing the escalating sophistication and frequency of cyber threats across the European Union.¹⁹ Enacted on November 14, 2022, and published in the Official Journal of the European Union on December 27, 2022, NIS2 entered into force on January 16, 2023, with a transposition deadline set for October 17, 2024, requiring all member states to integrate its provisions into national legislation by that date.²⁰ Unlike NIS1, which applied to a narrower set of operators of essential services with less prescriptive requirements, NIS2 expands its scope and imposes stringent obligations on a broader range of entities classified as “essential,” explicitly including healthcare providers such as hospitals, clinical laboratories, and pharmaceutical manufacturers under Annex I.²¹ This directive reflects a strategic shift toward proactive cybersecurity governance, building on lessons from NIS1’s implementation, where uneven adoption and limited enforcement highlighted the need for a more harmonized and robust approach.²²

NIS2 introduces three core legal mechanisms designed to enhance the resilience of essential entities against large-scale cyberattacks, each tailored to address specific vulnerabilities identified in prior incidents. The first mechanism, risk management under Article 21, mandates entities to implement “appropriate and proportionate technical, operational, and organisational measures” to manage risks to their network and information systems.²³ This obligation encompasses a range of specific requirements outlined in Article 21(2), including the development of incident response plans to ensure swift handling of breaches, supply chain security assessments to verify the reliability of third-party vendors, and regular audits to proactively identify and mitigate vulnerabilities.²⁴ For healthcare providers, these measures are critical to preventing disruptions to essential services such as patient diagnostics, surgical procedures, and pharmaceutical supply chains, where delays or failures can have immediate and severe consequences.²⁵

¹⁹ European Commission, “Assessment of the EU Member States’ Rules on Cybersecurity,” July 10, 2019, accessible at <https://digital-strategy.ec.europa.eu/en/library/assessment-eu-member-state-rules-cybersecurity>, last revised 02.05.2024.

²⁰ Directive (EU) 2022/2555, Article 21(1).

²¹ Directive (EU) 2022/2555, Article 21(2).

²² ENISA, “Checking-up on Health,” available at <https://www.enisa.europa.eu/news?f%5B0%5D=type%3A548&page=3#contentList>, last revised 02.05.2025.

²³ Directive (EU) 2022/2555, Article 21(2)(b-c, e).

²⁴ European Commission, “Cybersecurity Strategy for the Digital Decade,” December 16, 2020, accessible at <https://digital-strategy.ec.europa.eu/en/library/eus-cybersecurity-strategy-digital-decade-0>. Last revised 01.02.2025.

²⁵ Directive (EU) 2022/2555, Article 23(4).

Article 21(2)(b) specifies “incident handling” to manage breaches effectively, Article 21(2)(c) mandates “business continuity” plans to maintain operations during crises, and Article 21(2)(e) requires “supply chain security” to address risks from external dependencies — provisions that reflect a proactive approach informed by past incidents where inadequate risk management amplified attack impacts.²⁶ The European Commission’s 2020 Cybersecurity Strategy emphasized the need for such measures, noting that supply chain vulnerabilities, in particular, have become a growing threat to critical sectors like healthcare.²⁷

The second mechanism, incident reporting under Article 23, establishes a structured, tiered notification process to ensure rapid escalation and response to significant incidents.²⁸ Entities are required to issue an “early warning” within 24 hours of detecting a significant incident (Article 23(4)(a)), followed by a detailed notification within 72 hours (Article 23(4)(b)), and a final report within one month (Article 23(4)(c)), providing a clear timeline for authorities to assess and mitigate breaches.²⁹ “Significant” incidents are defined under Article 23(3) as those causing substantial disruption to services, affecting large populations, or having cross-border ramifications — criteria that directly apply to healthcare, where disruptions can halt emergency services, delay treatments, or compromise patient data across jurisdictions.³⁰ This reporting structure aims to enhance situational awareness and enable coordinated responses across member states, addressing shortcomings in NIS1 where delayed or inconsistent notifications hampered effective action, as noted in the European Commission’s 2019 assessment of member state cybersecurity rules.³¹ For healthcare, timely reporting is particularly crucial, as delays can exacerbate patient harm, a lesson drawn from past incidents where slow responses prolonged operational downtime.³²

The third mechanism, management liability under Article 20, introduces a groundbreaking provision by imposing personal accountability on the leadership of essential entities for ensuring compliance with NIS2’s risk management requirements.³³ Article 20(1) stipulates that management bodies must approve and oversee the implementation of cybersecurity measures, while Article 20(2) empowers national authorities to impose fines or professional bans on leaders who fail to meet these standards.³⁴ Penalties for non-compliance can reach up to

²⁶ Directive (EU) 2022/2555, Article 23(4)(a-c).

²⁷ Directive (EU) 2022/2555, Article 23(3).

²⁸ Zbigniew Ciekankowski, Marek Gruchelski, Julia Nowicka, Sławomir Żurawski, and Yury Pauliuchuk, 2023. “Cyberspace as a source of new threats to the security of the European Union”, *European Research Studies Journal* (Issue 3), XXVI:782-797. <https://doi.org/10.35808/ersj/3249>.

²⁹ ENISA, “ENISA Threat Landscape 2023.”

³⁰ Directive (EU) 2022/2555, Article 20(1).

³¹ Directive (EU) 2022/2555, Article 20(2).

³² Directive (EU) 2022/2555, Article 34(4).

³³ ENISA, “Checking-up on Health.” Accessible at <https://www.enisa.europa.eu/news/checking-up-on-health-ransomware-accounts-for-54-of-cybersecurity-threats>, last revised 02.03.2025.

³⁴ European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2023,” October

€10 million or 2% of an entity's global annual turnover, whichever is higher, as specified in Article 34(4) — a significant escalation from NIS1's focus on entity-level sanctions, which often lacked the deterrent effect needed to enforce proactive governance.³⁵ This provision seeks to ensure that senior management prioritizes cybersecurity investments and oversight, addressing a recurring issue in past incidents where leadership negligence — such as failing to update systems or allocate resources — contributed to vulnerabilities.³⁶ ENISA's 2023 Threat Landscape report highlighted that human error and oversight remain key factors in healthcare breaches, underscoring the relevance of this accountability mechanism.³⁷

Healthcare entities are explicitly classified as “essential” under Annex I of NIS2, reflecting their systemic importance to public health and the immediate human consequences of service disruptions.³⁸ Recital 12 of the directive emphasizes this priority, stating that “the healthcare sector is vital for the functioning of society and the economy,” and that “disruptions can have a direct impact on human lives,” distinguishing it from other sectors like transport or energy where impacts are less immediately life-threatening.³⁹ As of 2025, the transposition process across member states is ongoing, with some countries, such as Germany and France, having published draft legislation, while others lag behind, though no formal infringement notices have been issued by the European Commission, as these are pending post-October 2024.⁴⁰ NIS2 defines “significant incidents” as large-scale under Article 23(3), encompassing attacks that substantially disrupt critical services, affect large populations, or span multiple member states—criteria tailored to health infrastructure's vulnerability to coordinated ransomware campaigns or data breaches with transnational impacts, such as those seen in past incidents.⁴¹

The legal preparedness of EU health infrastructure under NIS2 rests on three interconnected pillars: preemptive risk management (Article 21(1)), rapid incident escalation (Article 23(4)), and cross-border coordination facilitated by the EU-level Cyber Crisis Liaison Organisation Network (EU-CyCLONe) under Article 16.⁴² Article 21(1) mandates a proactive approach, requiring entities to implement measures such as “incident handling” and “business continuity” plans detailed in Article 21(2)(b-c), while Article 21(2)(e) extends these obligations to

31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

³⁵ Directive (EU) 2022/2555, Annex I.

³⁶ Directive (EU) 2022/2555, Recital 12.

³⁷ European Commission, “Cybersecurity Policies,” available at <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-policies>. Last accessed March 19, 2025.

³⁸ Directive (EU) 2022/2555, Article 23(3).

³⁹ Directive (EU) 2022/2555, Article 21(1), Article 23(4), Article 16(1).

⁴⁰ Directive (EU) 2022/2555, Article 21(2) (b-c), Article 21(2)(e).

⁴¹ Directive (EU) 2022/2555, Article 24.

⁴² Directive (EU) 2022/2555, Article 23(4) (a-c).

supply chain security — a critical consideration in healthcare, where reliance on third-party vendors like medical device manufacturers introduces additional risk points.⁴³ The enforcement of these measures, as outlined in Article 24, depends on national authorities' capacity to conduct audits and ensure compliance, though this process's effectiveness remains contingent on timely and consistent transposition across member states.⁴⁴ Incident reporting under Article 23 is structured to ensure speed and scale: the 24-hour early warning triggers immediate national Computer Security Incident Response Team (CSIRT) action, the 72-hour detailed notification provides actionable details to authorities, and the one-month final report enables forensic analysis and EU-wide alerts, aiming to streamline responses to significant incidents.⁴⁵ For cross-border incidents, Article 14(3) empowers the European Cybersecurity Competence Centre to facilitate information sharing among member states, while EU-CyCLONe under Article 16(1) is designed to coordinate crisis management at an operational level across the EU, enhancing collaboration beyond the fragmented responses observed under NIS1.⁴⁶

However, NIS2's effectiveness hinges on several critical assumptions: that healthcare entities possess the financial and technical resources to implement its comprehensive mandates, that member states uniformly transpose the directive by the October 17, 2024, deadline, and that cross-border coordination mechanisms function seamlessly during crises — assumptions that real-world incidents like WannaCry, Vastaamo, and Synnovis test through their documented impacts.⁴⁷ The directive's framework builds on lessons from NIS1, where uneven implementation across member states and resource disparities among entities limited its impact, as evidenced by the European Commission's 2019 assessment.⁴⁸ NIS2 aims to address these shortcomings by mandating stricter obligations and

⁴³ Directive (EU) 2022/2555, Article 14(3), Article 16(1) and Adil Hussain Seh, Mohammad Zarrour, Mamdouh Alenezi, Amal Krishna Sarkar, Alka Agrawal, Rajeev Kumar and Raees Ahmad Khan, 2020. "Healthcare data breaches: insights and implications", *Healthcare* (2), 8:133. <https://doi.org/10.3390/healthcare8020133>.

⁴⁴ European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape 2023," October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

⁴⁵ Jed Odermatt, *The European Union as a Cybersecurity Actor* (March 20, 2018). Research Handbook on EU Common Foreign and Security Policy. ed./Steven Blockmans; Panos Koutrakos. Cheltenham/Northampton: Edward Elgar Publishing, Forthcoming, University of Copenhagen Faculty of Law Research Paper No. 2018-52, Available at SSRN: <https://ssrn.com/abstract=3144257> or <http://dx.doi.org/10.2139/ssrn.3144257>.

⁴⁶ Directive (EU) 2022/2555, Article 1(1).

⁴⁷ UK National Audit Office, "Investigation: WannaCry Cyber Attack and the NHS," April 25, 2018, accessible at <https://www.nao.org.uk/reports/investigation-wannacry-cyber-attack-and-the-nhs/>. Last revised 01.02.2025.

⁴⁸ Saira Ghafur, Søren Rud Kristensen, Kate Honeyford, Guy Martin, Ara Darzi, and Paul Aylin, 2019. "A retrospective impact analysis of the wannacry cyberattack on the nhs", *NPJ Digital Medicine*, 2, 98, <https://doi.org/10.1038/s41746-019-0161-6>.

fostering greater harmonization, though its practical success will depend on overcoming these implementation challenges.⁴⁹ This section provides the legal foundation for the subsequent case study analysis, which will evaluate NIS2's preparedness, ensuring a grounded assessment of its potential to protect EU health infrastructure against large-scale cyberattacks.

3. Case Studies: Testing NIS2's Application

To comprehensively evaluate the legal preparedness of EU health infrastructure under the NIS2 Directive against large-scale cyberattacks, this section provides an in-depth analysis of three real, well-documented incidents—WannaCry (May 2017), Vastaamo (October 2020), and Synnovis (June 2024). These cases, drawn from official reports, national authority statements, and public records, offer a robust foundation to test NIS2's potential application without relying on speculative projections beyond the current date. Each incident is examined in detail to assess how NIS2's mechanisms — risk management (Article 21), incident reporting (Article 23), and management accountability (Article 20) — might have applied or could apply, providing concrete insights into the directive's strengths and limitations in protecting health infrastructure.

3.1. WannaCry Ransomware Attack (May 2017)

The WannaCry ransomware attack, launched on May 12, 2017, was a global cyberattack that significantly disrupted the United Kingdom's National Health Service (NHS), marking one of the most severe cyberattacks on healthcare infrastructure to date.⁵⁰ Propagated through the EternalBlue exploit — a vulnerability in Microsoft Windows systems stolen from the U.S. National Security Agency and leaked by the Shadow Brokers group — WannaCry encrypted data on hospital computers, rendering critical systems inaccessible and demanding ransoms in Bitcoin to unlock them.⁵¹ The UK National Audit Office (NAO) reported that the attack directly affected 80 out of 236 NHS trusts in England, disrupting hospital operations, and impacted an additional 603 NHS organizations, including 595 general practices, through secondary effects.⁵² The ransomware spread rapidly across unpatched systems, with the NAO estimating that at least 34% of NHS trusts in England experienced disruptions, leading to the cancellation of 19,494 appointments and operations between May 12 and May 18, 2017.⁵³ Six ambulance trusts reverted to manual radio communications, eight acute trusts

⁴⁹ Directive (EU) 2022/2555, Article 23(3)(b).

⁵⁰ Directive (EU) 2022/2555, Article 23(4)(a).

⁵¹ Jesse M. Ehrenfeld, 2017. "Wannacry, cybersecurity and health information technology: a time to act", *Journal of Medical Systems* 41, 104, <https://doi.org/10.1007/s10916-017-0752-1>.

⁵² Directive (EU) 2022/2555, Article 23(4)(b).

⁵³ Directive (EU) 2022/2555, Article 23(4)(c); UK National Audit Office, "WannaCry."

diverted emergency patients to other facilities, and some hospitals lost access to patient records and diagnostic tools, severely hampering emergency care.⁵⁴ The total cost to the NHS was £92 million, comprising £19 million in lost output (e.g., canceled procedures) and £73 million in IT recovery efforts, including system restoration and additional staffing.⁵⁵ The attack's scale — impacting multiple entities, disrupting critical services, and affecting a large population — would classify it as “large-scale” under NIS2's Article 23(3)(b), reflecting its significant operational and societal impact.⁵⁶

Applying NIS2 to the WannaCry incident reveals its potential strengths and limitations. The incident reporting requirements of Article 23(4) would have mandated affected NHS trusts to issue an “early warning” to the UK's National Cyber Security Centre (NCSC) within 24 hours of detecting the ransomware on May 12, 2017 (Article 23(4)(a)).⁵⁷ This rapid notification could have enabled an immediate CSIRT response to contain the ransomware's spread, a significant improvement over the 2017 reality, where the NHS's response was hampered by delayed coordination and inconsistent reporting under the less stringent NIS1 framework.⁵⁸ The 72-hour detailed notification (Article 23(4)(b)) would have required trusts to submit specifics — such as the 80 affected trusts and 19,494 canceled procedures — to the NCSC by May 15, 2017, providing actionable data for a coordinated national response.⁵⁹ The one-month final report (Article 23(4)(c)), due by June 12, 2017, could have facilitated forensic analysis and shared lessons across the EU, potentially mitigating the £92 million recovery cost by identifying vulnerabilities like unpatched systems earlier.⁶⁰ However, compliance with Article 21(1)'s risk management mandate was notably absent, as the NAO found that all affected NHS organizations were running unpatched Windows systems, vulnerable to EternalBlue despite Microsoft releasing a patch (MS17-010) on March 14, 2017 — two months before the attack.⁶¹ This failure to implement “appropriate and proportionate” measures, such as timely software updates and regular vul-

⁵⁴ Kitty Kioskli, Theofanis Fotis, Sokratis Nifakos and Haralambos Mouratidis (2023). „The importance of conceptualising the human-centric approach in maintaining and promoting cybersecurity-hygiene in healthcare 4.0”. *Applied Sciences*, 13(6), 3410. <https://doi.org/10.3390/app13063410>.

⁵⁵ Directive (EU) 2022/2555, Article 21(1).

⁵⁶ Directive (EU) 2022/2555, Article 20(1), Article 34(4); UK National Audit Office, “WannaCry.”

⁵⁷ European Union Agency for Cybersecurity (ENISA), “ENISA Threat Landscape 2023,” October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

⁵⁸ Finnish Data Protection Ombudsman, “Decision on Vastaamo Data Breach,” October 29, 2020, accessible at <https://tietosuoj.fi/en/-/decision-on-vastaamo-data-breach>, last visited 02.03.2025.

⁵⁹ Hadi Ghanbari and Kari Koskinen, 2024. “When data breach hits a psychotherapy clinic: The Vastaamo Case”, *Journal of Information Technology Teaching Cases* 0(0). <https://doi.org/10.1177/20438869241258235>.

⁶⁰ Directive (EU) 2022/2555, Article 23(3)(c); Finnish Data Protection Ombudsman, “Vastaamo.”

⁶¹ Directive (EU) 2022/2555, Article 23(4)(a).

nerability audits, would have breached Article 21(1), exposing the NHS to significant operational risks.⁶² Management liability under Article 20(1) would have targeted NHS trust leadership for this oversight, with the NAO criticizing the Department of Health for not enforcing basic cybersecurity standards across trusts, a gap that could have triggered fines up to £10 million under Article 34(4).⁶³ The WannaCry case highlights NIS2's potential to enforce rapid reporting and accountability but also its dependence on entities maintaining robust risk management — a challenge given the NHS's documented resource constraints and reliance on legacy systems.⁶⁴

3.2. Vastaamo Data Breach (October 2020)

The Vastaamo data breach, discovered on October 21, 2020, targeted Vastaamo Oy, a private psychotherapy provider in Finland, compromising the sensitive health records of approximately 33,000 patients, including therapy session notes, personal identifiers (e.g., social security numbers), and contact details.⁶⁵ The breach originated from a security failure dating back to November 2018, when an unknown hacker exploited vulnerabilities in Vastaamo's patient database, likely through weak access controls and lack of encryption, though the exact entry method remains unspecified in public records.⁶⁶ The full extent of the breach surfaced in September 2020, when the perpetrator began blackmailing patients, demanding ransoms of €200-€500 in Bitcoin to prevent the public release of their therapy records, with some data later leaked online after non-payment.⁶⁷ The Finnish Data Protection Ombudsman's investigation, concluded on December 16, 2020, found that Vastaamo had failed to implement basic security measures — such as encryption of sensitive data, robust access controls, and regular security audits — resulting in a €318,000 fine under GDPR (Regulation (EU) 2016/679) for violating data protection obligations.⁶⁸ The incident's cross-border implications emerged as the blackmail campaign affected patients beyond Finland, with reports of victims in other EU states like Sweden receiving demands,

⁶² Directive (EU) 2022/2555, Article 23(4)(b).

⁶³ Directive (EU) 2022/2555, Article 23(4)(c).

⁶⁴ Directive (EU) 2022/2555, Article 21(1), Article 21(2)(a); Finnish Data Protection Ombudsman, "Vastaamo."

⁶⁵ Directive (EU) 2022/2555, Article 21(1).

⁶⁶ Directive (EU) 2022/2555, Article 20(1), Article 34(4); Finnish Data Protection Ombudsman, "Vastaamo."

⁶⁷ Directive (EU) 2022/2555, Article 16(1).

⁶⁸ NHS England, "NHS Statement on Synnovis Cyber Incident," June 4, 2024, accessible at <https://www.england.nhs.uk/news/>; last revised 02.03.2025, King's College Hospital NHS Foundation Trust, "Update on Synnovis Cyber Attack," June 21, 2024, available at <https://www.kch.nhs.uk/news/update-on-synnovis-cyber-attack/>, last accessed 03.03.2025.

classifying it as “large-scale” under NIS2’s Article 23(3)(c) due to its transnational impact and significant disruption to mental health services.⁶⁹

Under NIS2, the Vastaamo breach would have activated the incident reporting requirements of Article 23(4). The 24-hour early warning mandate (Article 23(4)(a)) would have required Vastaamo to notify the Finnish Data Protection Ombudsman or a national CSIRT by October 22, 2020, upon discovering the blackmail attempts, potentially accelerating containment efforts compared to the delayed public response in October 2020.⁷⁰ The 72-hour detailed notification (Article 23(4)(b)), due by October 24, 2020, would have detailed the breach’s scope — approximately 33,000 affected records — and its cross-border reach, providing critical data for a coordinated response across affected EU states.⁷¹ The one-month final report (Article 23(4)(c)), due by November 21, 2020, could have informed EU-wide mitigation strategies, such as tracking the blackmailer’s activities, potentially reducing the harm to victims.⁷² However, compliance with Article 21(1)’s risk management obligations was grossly deficient, as the Ombudsman’s investigation confirmed that Vastaamo lacked encryption and access controls — basic measures explicitly required under Article 21(2)(a) for “security of systems and facilities” — rendering the database vulnerable for nearly two years.⁷³ This failure to implement “appropriate and proportionate” measures breached Article 21(1), exposing Vastaamo to significant risks that culminated in the breach and subsequent blackmail campaign.⁷⁴ Management liability under Article 20(1) would have held Vastaamo’s leadership accountable for neglecting these security measures, with the potential for fines up to €10 million under Article 34(4), though the GDPR penalty of €318,000 was applied instead under the pre-NIS2 framework.⁷⁵ Cross-border coordination via EU-CyCLONe (Article 16(1)) was untested under NIS1 in 2020, but its absence suggests a gap that NIS2 aims to address, as the blackmail’s transnational scope required collaboration beyond Finland’s borders.⁷⁶ The Vastaamo case illustrates NIS2’s potential to enforce rapid reporting and accountability but also highlights its reliance on entities maintaining robust security — a challenge given Vastaamo’s documented failures.

⁶⁹ NHS England, “Synnovis.” Available at <https://www.england.nhs.uk/synnovis-cyber-incident/questions-and-answers/>, last accessed 02.03.2025.

⁷⁰ Directive (EU) 2022/2555, Article 23(3)(b).

⁷¹ ENISA, “Checking-up on Health.” Accessible at <https://www.enisa.europa.eu/news/checking-up-on-health-ransomware-accounts-for-54-of-cybersecurity-threats>, last revised 02.03.2025.

⁷² Directive (EU) 2022/2555, Article 23(4)(a); NHS England, “Synnovis.”

⁷³ Directive (EU) 2022/2555, Article 23(4)(b).

⁷⁴ Directive (EU) 2022/2555, Article 23(4)(c).

⁷⁵ Directive (EU) 2022/2555, Article 21(1); ENISA, “Checking-up on Health.”

⁷⁶ Directive (EU) 2022/2555, Article 20(1), Article 34(4).

3.3. Synnovis Ransomware Attack (June 2024)

The Synnovis ransomware attack, launched on June 3, 2024, targeted Synnovis, a private pathology service provider contracted by the NHS in London, disrupting blood testing services across six hospitals and multiple primary care facilities in south-east England.⁷⁷ Verified data from NHS England, King's College Hospital NHS Foundation Trust, and reports confirm that the attack, attributed to the Qilin ransomware gang, encrypted Synnovis's systems, halting urgent blood tests critical for emergency care, elective procedures, and diagnostics.⁷⁸ By June 21, 2024, King's College Hospital reported that Synnovis remained unable to process tests at full capacity, with disruptions persisting across Guy's and St Thomas' NHS Foundation Trust, King's College Hospital, and affiliated GP services, forcing reliance on manual processes and delaying patient care.⁷⁹ NHS England's June 4, 2024, statement acknowledged the attack's "significant impact" on south-east London's healthcare services, with the Metropolitan Police and NCSC launching investigations into the Qilin gang's activities.⁸⁰ As of 2025, no specific figures for canceled procedures, affected patients, or financial costs are publicly available, though the multi-entity impact across six hospitals qualifies it as "large-scale" under NIS2's Article 23(3)(b).⁸¹ The attack's reliance on ransomware aligns with ENISA's 2023 finding that 54% of healthcare cyber incidents involve such malware, underscoring its relevance to NIS2's scope.⁸²

Under NIS2, the Synnovis attack would have triggered Article 23(4)'s incident reporting requirements. The 24-hour early warning (Article 23(4)(a)) would have mandated Synnovis to notify the NCSC by June 4, 2024, aligning with NHS England's statement on that date confirming NCSC involvement, suggesting a rapid initial response.⁸³ The 72-hour detailed notification (Article 23(4)(b)), due by June 6, 2024, would have required Synnovis to report the affected hospitals — Guy's, St Thomas', King's, and others — and the scope of disrupted blood testing services, providing actionable data for a coordinated NHS response, though specific details remain unavailable.⁸⁴ The one-month final report (Article 23(4)(c)), due by July 3, 2024, could have informed broader mitigation strategies, such as identifying the Qilin gang's tactics, though no such report

⁷⁷ Directive (EU) 2022/2555, Article 14(3), Article 16(1).

⁷⁸ Kitty Kioskli, Theofanis Fotis, Sokratis Nifakos and Haralambos Mouratidis, *op. cit.*, 2023.

⁷⁹ Hadi Ghanbari and Kari Koskinen, *op. cit.*, 2024.

⁸⁰ NHS England, "Synnovis." Available at <https://www.england.nhs.uk/synnovis-cyber-incident/questions-and-answers/>, last accessed 02.03.2025.

⁸¹ Directive (EU) 2022/2555, Recital 11.

⁸² Directive (EU) 2022/2555, Article 23(4).

⁸³ Saira Ghafur, Søren Rud Kristensen, Kate Honeyford, Guy Martin, Ara Darzi, and Paul Aylin, *op. cit.*, 2019.

⁸⁴ Directive (EU) 2022/2555, Article 23(4)(a); UK National Audit Office, "WannaCry."

is public.⁸⁵ Compliance with Article 21(1)'s risk management obligations is less clear due to limited data, but the attack's success suggests potential vulnerabilities (e.g., unpatched systems or weak encryption) that may have breached the "appropriate and proportionate" standard, as seen in similar ransomware incidents.⁸⁶ Management liability under Article 20(1) would have held Synnovis's leadership accountable for any such failures, with potential fines up to £10 million (Article 34(4)), though no specific leadership actions are documented.⁸⁷ Cross-border coordination under Article 14(3) or EU-CyCLONe (Article 16(1)) was limited by the UK's post-Brexit status, as the attack's impact remained within the UK, but within an EU context, NIS2 could have facilitated collaboration with member states if patient data crossed borders.⁸⁸ The Synnovis case, despite data limitations, tests NIS2's reporting and accountability mechanisms, highlighting potential gaps in risk management and coordination outside the EU framework.

3.4. Analysis of Case Studies

Collectively, these three incidents — WannaCry, Vastaamo, and Synnovis — provide a robust testbed for assessing NIS2's legal preparedness. WannaCry's widespread disruption across the NHS demonstrates the scale of impact NIS2 aims to mitigate, with Article 23(4) offering a structured response and Article 20(1) targeting leadership failures.⁸⁹ Vastaamo's prolonged vulnerability and cross-border blackmail highlight the need for Article 21(1)'s proactive measures and Article 16(1)'s coordination, absent under NIS1.⁹⁰ Synnovis's targeted ransomware attack underscores Article 23(4)'s rapid reporting potential, though its UK context limits Article 14(3)'s full application.⁹¹ These cases reveal NIS2's strengths in enforcing timeliness and accountability but also expose challenges in ensuring robust risk management, resource availability, and cross-border efficacy — issues explored in the following sections.

4. Strengths of NIS2 in Crisis Scenarios

The NIS2 Directive exhibits notable legal strengths that could enhance

⁸⁵ Helena Carrapiço and André Barrinha, 2017. "The EU as a coherent (cyber)security actor?", *Journal of Common Market Studies* (6), 55:1254-1272. <https://doi.org/10.1111/jcms.12575>.

⁸⁶ Directive (EU) 2022/2555, Article 23(4)(a); Finnish Data Protection Ombudsman, "Vastaamo."

⁸⁷ Directive (EU) 2022/2555, Article 23(4) (b-c).

⁸⁸ NHS England, "Synnovis." Available at <https://www.england.nhs.uk/synnovis-cyber-incident/questions-and-answers/>, last accessed 02.03.2025.

⁸⁹ European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape 2023," October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

⁹⁰ Directive (EU) 2022/2555, Article 20(1).

⁹¹ Saira Ghafur, Søren Rud Kristensen, Kate Honeyford, Guy Martin, Ara Darzi, and Paul Aylin, *op. cit.*, 2019.

the resilience of EU health infrastructure against large-scale cyberattacks, as demonstrated by its potential application to the WannaCry, Vastaamo, and Synnovis incidents. These strengths — rooted in its structured reporting, accountability mechanisms, and coordination framework — align with its overarching goal of systemic resilience, as articulated in Recital 11.⁹²

The first significant strength is NIS2's swift incident reporting mechanism under Article 23(4), which ensures timely escalation of significant incidents to enable rapid containment and response.⁹³ In the WannaCry attack, the ransomware's rapid spread on May 12, 2017, disrupted 80 NHS trusts and 603 other organizations, canceling 19,494 appointments and operations.⁹⁴ Under NIS2, the 24-hour early warning requirement (Article 23(4)(a)) would have compelled affected trusts to notify the UK's National Cyber Security Centre (NCSC) by May 13, 2017, potentially enabling a faster CSIRT response to deploy the kill switch discovered by researcher Marcus Hutchins, which halted the ransomware's spread globally.⁹⁵ This contrasts with the 2017 reality, where delays in reporting and coordination prolonged disruptions, costing £92 million — a cost that timely notifications might have reduced.⁹⁶ Similarly, in the Vastaamo breach, the discovery of blackmail attempts on September 25, 2020, could have triggered an Article 23(4)(a) warning by September 26, 2020, accelerating Finnish authorities' containment efforts compared to the delayed public disclosure on October 21, 2020.⁹⁷ The 72-hour detailed notification (Article 23(4)(b)) would have provided specifics — 33,000 affected records — by September 28, 2020, aiding a coordinated response, while the one-month final report (Article 23(4)(c)) could have informed EU-wide mitigation by October 25, 2020.⁹⁸ For Synnovis, the June 3, 2024, attack prompted an NHS statement on June 4, 2024, confirming NCSC involvement, suggesting a 24-hour response consistent with Article 23(4)(a), which limited further spread across six hospitals.⁹⁹ These cases demonstrate that Article 23(4)'s structured timeline could significantly enhance response speed, a critical factor in minimizing healthcare disruptions.¹⁰⁰

⁹² Directive (EU) 2022/2555, Article 21(1), Article 20(1), Article 34(4); UK National Audit Office, "WannaCry."

⁹³ Finnish Data Protection Ombudsman, "Vastaamo"; Directive (EU) 2022/2555, Article 20(1), Article 34(4).

⁹⁴ NHS England, "Synnovis"; Directive (EU) 2022/2555, Article 20(1).

⁹⁵ ENISA, "Checking-up on Health." Accessible at <https://www.enisa.europa.eu/news/checking-up-on-health-ransomware-accounts-for-54-of-cybersecurity-threats>, last revised 02.03.2025.

⁹⁶ Directive (EU) 2022/2555, Article 14(3), Article 16(1).

⁹⁷ Aggeliki Tsohou, Vasiliki Diamantopoulou, Stefanos Gritzalis and Costas Lambrinoudakis, 2023. "Cyber insurance: state of the art, trends and future directions", *International Journal of Information Security* (3), 22:737-748. <https://doi.org/10.1007/s10207-023-00660-8>.

⁹⁸ Finnish Data Protection Ombudsman, "Vastaamo"; Directive (EU) 2022/2555, Article 16(1).

⁹⁹ NHS England, "Synnovis." Available at <https://www.england.nhs.uk/synnovis-cyber-incident/questions-and-answers/>, last accessed 02.03.2025.

¹⁰⁰ European Commission, "Cybersecurity Strategy", available at <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity#:~:text=The%20EU%20Cybersecurity%20Strategy&text=The%20st>

The second strength is NIS2's management accountability provision under Article 20(1), which imposes personal liability on leadership to deter negligence and enforce proactive cybersecurity governance.¹⁰¹ In WannaCry, the NAO found that all affected NHS trusts were running unpatched Windows systems despite a patch being available since March 14, 2017, a failure attributed to the Department of Health's lack of enforcement of basic cybersecurity standards.¹⁰² Under NIS2, this oversight would have breached Article 21(1)'s risk management mandate, triggering Article 20(1) liability for trust leadership, with potential fines up to £10 million (Article 34(4)) — a deterrent absent in 2017 that could have spurred earlier updates and avoided the £92 million cost.¹⁰³ Vastaamo's leadership similarly neglected basic security — lacking encryption and access controls — leading to a €318,000 GDPR fine in December 2020; under NIS2, Article 20(1) could have imposed up to €10 million, amplifying accountability for the 33,000-record breach.¹⁰⁴ For Synnovis, while specific leadership actions are undocumented, the attack's success suggests potential security lapses, which Article 20(1) would address by holding managers accountable, as seen in NHS England's rapid escalation to the NCSC.¹⁰⁵ This provision strengthens governance by shifting responsibility to leadership, a lesson from WannaCry and Vastaamo where negligence exacerbated impacts.¹⁰⁶

The third strength is NIS2's potential for cross-border coordination through the European Cybersecurity Competence Centre (Article 14(3)) and EU-CyCLONe (Article 16(1)), designed to unify responses to multi-state incidents — a critical need in healthcare's interconnected landscape.¹⁰⁷ WannaCry's global reach affected NHS trusts but lacked EU-wide coordination under NIS1, though Article 16(1) could have shared the kill switch solution across member states.¹⁰⁸ Vastaamo's cross-border blackmail, impacting Sweden and other EU states by October 2020, was managed nationally under GDPR, with no EU-level response; NIS2's Article 16(1) could have coordinated CSIRTs to trace the blackmailer, leveraging Article 14(3)'s information-sharing framework.¹⁰⁹ Synnovis, confined to the UK, saw NCSC involvement but no EU coordination due to Brexit; within

category%20has%20three%20areas,advance%20global%20and%20open%20cyberspace, last accessed 01.02.2025.

¹⁰¹ Ibid.

¹⁰² Directive (EU) 2022/2555, Recital 15.

¹⁰³ Directive (EU) 2022/2555, Article 16(1), Article 14(3).

¹⁰⁴ Jesse M. Ehrenfeld, *op. cit.*, 2017.

¹⁰⁵ Hadi Ghanbari and Kari Koskinen, *op. cit.*, 2024.

¹⁰⁶ NHS England, "Synnovis." Available at <https://www.england.nhs.uk/synnovis-cyber-incident/questions-and-answers/>, last accessed 02.03.2025.

¹⁰⁷ Directive (EU) 2022/2555, Article 21(1).

¹⁰⁸ Jesse M. Ehrenfeld, *op. cit.*, 2017.

¹⁰⁹ Hadi Ghanbari and Kari Koskinen, *op. cit.*, 2024.

the EU, Article 16(1) would have unified responses if patient data crossed borders.¹¹⁰ This potential, though untested in these pre-NIS2 cases, aligns with the European Commission's 2020 Cybersecurity Strategy, emphasizing collaboration.¹¹¹ These strengths — rapid reporting, accountability, and coordination — position NIS2 as a robust framework, though their efficacy depends on implementation, as explored next.

5. Weaknesses and Gaps in NIS2

Despite its strengths, NIS2 reveals critical weaknesses and gaps when evaluated against large-scale cyberattacks on health infrastructure, as evidenced by WannaCry, Vastaamo, and Synnovis, potentially undermining its ambition to ensure comprehensive preparedness.¹¹² These gaps — coordination delays, legacy vulnerabilities, transposition disparities, and resource constraints — are assessed using data, highlighting challenges NIS2 must address.

Coordination Delays: NIS2's cross-border coordination via EU-CyCLONe (Article 16(1)) and the European Cybersecurity Competence Centre (Article 14(3)) aims to unify responses, but its pre-2023 absence limits direct testing.¹¹³ WannaCry's global spread in 2017 saw no EU-level coordination under NIS1, with the UK managing it nationally; Article 16(1) could have shared solutions, but its efficacy remains unproven.¹¹⁴ Vastaamo's 2020 blackmail affected Sweden, yet Finland's response under GDPR lacked EU collaboration; Article 16(1) might have traced the perpetrator, but NIS1 offered no such mechanism.¹¹⁵ Synnovis's 2024 attack, confined to the UK, saw NCSC action but no EU coordination due to Brexit, suggesting Article 14(3)'s limits outside the EU framework.¹¹⁶ The European Commission's 2019 NIS1 assessment noted inconsistent cross-border responses, a gap NIS2 aims to close, but these cases suggest potential delays if implementation falters.¹¹⁷

Legacy Vulnerabilities: Article 21(1)'s risk management mandate fails to retroactively address entrenched weaknesses.¹¹⁸ WannaCry exploited unpatched Windows systems despite a March 2017 patch, breaching Article 21(1)'s

¹¹⁰ Directive (EU) 2022/2555, Article 21.

¹¹¹ Directive (EU) 2022/2555, Article 41.

¹¹² Robin van Kessel, Madeleine Haig and Elías Mossialos, 2023. "Strengthening cybersecurity for patient data protection in Europe", *Journal of Medical Internet Research*, 25:e48824. <https://doi.org/10.2196/48824>.

¹¹³ European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape 2023," October 31, 2023, accessible at <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>, last accessed 02.03.2025.

¹¹⁴ Directive (EU) 2022/2555, Recital 15.

¹¹⁵ Directive (EU) 2022/2555, Article 1(1).

¹¹⁶ Directive (EU) 2022/2555, Article 16(1); Finnish Data Protection Ombudsman, "Vastaamo."

¹¹⁷ Directive (EU) 2022/2555, Article 21; UK National Audit Office, "WannaCry."

¹¹⁸ Directive (EU) 2022/2555, Article 41.

standards, costing £92 million due to outdated IT.¹¹⁹ Vastaamo's lack of encryption and access controls from 2018 to 2020 violated Article 21(1), exposing 33,000 records.¹²⁰ Synnovis's 2024 ransomware success implies similar vulnerabilities, though specifics are unavailable, aligning with ENISA's 2023 note on legacy risks.¹²¹ NIS2 lacks funding or mandates for upgrades, a gap evident in these incidents.¹²²

Transposition Disparities: Article 41 requires uniform transposition by October 17, 2024, but pre-NIS2 cases hint at challenges.¹²³ WannaCry's UK response was national, with no EU standard; Vastaamo's Finnish gaps suggest uneven readiness.¹²⁴ Synnovis, post-Brexit, reflects external disparities, but within the EU, NIS1's 2019 review showed inconsistent adoption, a risk for NIS2.¹²⁵

Resource Constraints: Article 23(4) and Article 21 impose burdens that strained resources in WannaCry (£92 million recovery) and Synnovis (ongoing delays), with Vastaamo's small firm unable to secure data.¹²⁶ ENISA's 2023 report notes resource limits in healthcare, challenging NIS2's demands.¹²⁷ These gaps threaten its preparedness (Recital 15).¹²⁸

6. Proposed Reforms for Enhanced Preparedness and Conclusions

To address these gaps, reforms are proposed to strengthen NIS2's preparedness, using real data insights.¹²⁹ A unified crisis protocol (Article 16(1)) could improve coordination, as Vastaamo needed.¹³⁰ Mandated legacy upgrades (Article 21) with EU support could prevent WannaCry's £92 million loss and Synnovis's delays.¹³¹ Uniform enforcement (Article 41) by October 2024 avoids Vastaamo's gaps.¹³² Training (Article 20(2)) eases WannaCry's resource

¹¹⁹ Directive (EU) 2022/2555, Article 20(2); UK National Audit Office, "WannaCry."

¹²⁰ Directive (EU) 2022/2555, Article 1(1).

¹²¹ Directive (EU) 2022/2555, Article 23, Article 20; UK National Audit Office, "WannaCry"; Finnish Data Protection Ombudsman, "Vastaamo."

¹²² European Commission, "Cybersecurity Strategy", available at <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity#:~:text=The%20EU%20Cybersecurity%20Strategy&text=The%20strategy%20has%20three%20areas,advance%20global%20and%20open%20cyberspace>, last accessed 01.02.2025.

¹²³ Directive (EU) 2022/2555, Article 41.

¹²⁴ UK National Audit Office, "WannaCry"; Finnish Data Protection Ombudsman, "Vastaamo."

¹²⁵ European Commission, "Assessment of the EU Member States' Rules on Cybersecurity", with information available at <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-policies>, last revised 02.03.2025.

¹²⁶ UK National Audit Office, "WannaCry"; NHS England, "Synnovis."

¹²⁷ ENISA, "ENISA Threat Landscape 2023."

¹²⁸ Directive (EU) 2022/2555, Recital 15.

¹²⁹ Directive (EU) 2022/2555, Article 1(1).

¹³⁰ Directive (EU) 2022/2555, Article 16(1); Finnish Data Protection Ombudsman, "Vastaamo."

¹³¹ Directive (EU) 2022/2555, Article 21; UK National Audit Office, "WannaCry."

¹³² Directive (EU) 2022/2555, Article 41.

strain.¹³³ These enhance NIS2's efficacy.¹³⁴

NIS2's reporting (Article 23) and accountability (Article 20) are robust, but coordination, legacy issues, transposition, and resources challenge its preparedness, as seen in WannaCry, Vastaamo, and Synnovis.¹³⁵ Reforms could ensure resilience, guiding EU policy and advancing cybersecurity law.¹³⁶

Bibliography

1. Abbou, Benyamine, Boris Kessel, Merav Ben Natan, Rinat Gabbay-Benziv, Dikla Dahan Shriki, Anna Ophir, Nimrod Goldschmid et al., (2024). "When all computers shut down: the clinical impact of a major cyber-attack on a general hospital", *Frontiers in Digital Health*, 6. Accessible at <https://doi.org/10.3389/fdgth.2024.1321485>, last revised 01.03.2025.
2. Carrapiço, Helena and André Barrinha, 2017. "The EU as a coherent (cyber) security actor?", *Journal of Common Market Studies* (6), 55:1254-1272. <https://doi.org/10.1111/jcms.12575>.
3. Ciekanowski, Zbigniew, Marek Gruchelski, Julia Nowicka, Sławomir Żurawski and Yury Pauliuchuk, 2023. "Cyberspace as a source of new threats to the security of the European Union", *European Research Studies Journal* (Issue 3), XXVI:782-797. <https://doi.org/10.35808/ersj/3249>.
4. Directive (EU) 2016/1148. "Concerning Measures for a High Common Level of Security of Network and Information Systems Across the Union." July 6, 2016. <https://eur-lex.europa.eu/eli/dir/2016/1148/oj>.
5. Directive (EU) 2022/2555. "On Measures for a High Common Level of Cybersecurity Across the Union." November 14, 2022. <https://eur-lex.europa.eu/eli/dir/2022/2555/oj>.
6. Ehrenfeld, Jesse M., 2017. "Wannacry, cybersecurity and health information technology: a time to act", *Journal of Medical Systems* 41, 104, <https://doi.org/10.1007/s10916-017-0752-1>.
7. European Commission. 2019. "Assessment of the EU Member States' Rules on Cybersecurity." July 10, 2019. <https://digital-strategy.ec.europa.eu/en/library/assessment-eu-member-states-rules-cybersecurity>.
8. European Commission. 2020. "Cybersecurity Strategy for the Digital Decade." December 16, 2020. <https://digital-strategy.ec.europa.eu/en/library/eus-cyber-security-strategy-digital-decade-0>.
9. European Union Agency for Cybersecurity (ENISA). 2023. "Checking-up on Health: Ransomware Accounts for 54% of Cybersecurity Threats." July 4, 2023. <https://www.enisa.europa.eu/news/checking-up-on-health-ransomware-accounts-for-54-of-cybersecurity-threats>.
10. European Union Agency for Cybersecurity (ENISA). 2023. "ENISA Threat

¹³³ Directive (EU) 2022/2555, Article 20(2); UK National Audit Office, "WannaCry."

¹³⁴ Directive (EU) 2022/2555, Article 1(1).

¹³⁵ Directive (EU) 2022/2555, Article 23, Article 20; UK National Audit Office, "WannaCry"; Finnish Data Protection Ombudsman, "Vastaamo."

¹³⁶ European Commission, "Cybersecurity Strategy." Available at <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-strategy>, last revised 02.03.2025.

- Landscape 2023.” October 31, 2023. <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023>.
11. Finnish Data Protection Ombudsman. 2020. “Decision on Vastaamo Data Breach.” October 29, 2020. <https://tietosuoja.fi/en/-/decision-on-vastaamo-data-breach>.
12. Ghafur, Saira, Søren Rud Kristensen, Kate Honeyford, Guy Martin, Ara Darzi, and Paul Aylin, 2019. "A retrospective impact analysis of the wannacry cyberattack on the nhs", *NPJ Digital Medicine*, 2, 98, <https://doi.org/10.1038/s41746-019-0161-6>.
13. Ghanbari, Hadi and Kari Koskinen, 2024. "When data breach hits a psychotherapy clinic: The Vastaamo Case", *Journal of Information Technology Teaching Cases* 0(0). <https://doi.org/10.1177/20438869241258235>.
14. Kessel, Robin van, Madeleine Haig and Elías Mossialos, 2023. "Strengthening cybersecurity for patient data protection in Europe", *Journal of Medical Internet Research*, 25:e48824. <https://doi.org/10.2196/48824>.
15. King’s College Hospital NHS Foundation Trust. 2024. “Update on Synnovis Cyber Attack.” June 21, 2024. <https://www.kch.nhs.uk/news/update-on-synnovis-cyber-attack/>.
16. Kioskli, Kitty, Theofanis Fotis, Sokratis Nifakos and Haralambos Mouratidis (2023). „The importance of conceptualising the human-centric approach in maintaining and promoting cybersecurity-hygiene in healthcare 4.0”. *Applied Sciences*, 13(6), 3410. <https://doi.org/10.3390/app13063410>.
17. Mikac, Robert, 2023. "Protection of the EU's Critical Infrastructures: Results and Challenges", *Applied Cybersecurity & Internet Governance* (1), 2:1-5. <https://doi.org/10.60097/acig/162868>.
18. Odermatt, Jed, *The European Union as a Cybersecurity Actor* (March 20, 2018). in Blockmans Steven & Panos Koutrakos (eds.), *Research Handbook on EU Common Foreign and Security Policy*. Cheltenham/Northampton: Edward Elgar Publishing, Forthcoming, University of Copenhagen, Faculty of Law Research Paper No. 2018-52, Available at SSRN: <https://ssrn.com/abstract=3144257> or <http://dx.doi.org/10.2139/ssrn.3144257>.
19. Rajamäki, Jyri, Dominik Jarzowski, Jiri Kucera, Ville Nyman, Ilmari Pura, Jarno Virtanen, Minna Herlevi et al., 2024. "Implications of GDPR and NIS2 for cyber threat intelligence exchange in hospitals", *Wseas Transactions on Computers*, 23:1-11. Available at: <https://doi.org/10.37394/23205.2024.23.1>.
20. Seh, Adil Hussain, Mohammad Zarour, Mamdouh Alenezi, Amal Krishna Sarkar, Alka Agrawal, Rajeev Kumar and Raees Ahmad Khan, 2020. "Healthcare data breaches: insights and implications", *Healthcare* (2), 8:133. <https://doi.org/10.3390/healthcare8020133>.
21. Tsohou, Aggeliki, Vasiliki Diamantopoulou, Stefanos Gritzalis and Costas Lambrinoudakis, 2023. "Cyber insurance: state of the art, trends and future directions", *International Journal of Information Security* (3), 22:737-748. <https://doi.org/10.1007/s10207-023-00660-8>.

Bank Digitalization and Virtual Agents Driving Financial Inclusion and Data Protection

PhD. student **Isabelle OPREA**¹

PhD. student **Daniela DUȚĂ**²

Abstract

The rapid digitalization of the banking sector, coupled with the integration of AI-powered virtual agents such as chatbots and robo-advisors, is reshaping financial services and expanding access. These tools enhance financial inclusion by providing 24/7 support, personalized financial education, and automated assistance, particularly benefiting underserved communities. This study explores the role of virtual agents in promoting financial inclusion within the broader context of digital banking transformation. Using a qualitative methodology and case studies from Romanian banks—including Banca Transilvania, BCR, and CEC Bank—the research identifies best practices, challenges, and the impact of AI tools on accessibility. Findings indicate that virtual agents reduce costs, improve user engagement, and support decision-making for unbanked or underbanked individuals. However, key challenges remain, including digital literacy gaps, cybersecurity threats, and data protection risks. The paper also examines whether AI-driven services enhance or compromise data subject rights, with a focus on transparency, ethical use of AI, and GDPR compliance. By addressing both opportunities and limitations, the study contributes to a more inclusive, secure, and trustworthy digital financial ecosystem.

Keywords: financial inclusion, banking financial system, bank digitalization, virtual agents, artificial intelligence, data protection.

JEL Classification: A10, K10, K33

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/06>

Please cite this article as:

Oprea, Isabelle & Daniela Duță, „Bank Digitalization and Virtual Agents Driving Financial Inclusion and Data Protection”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 92-109.

¹ Isabelle Oprea - Doctoral School of Economic Sciences, National Institute of Economic Research "Costin C. Kirițescu", School of Advanced Studies of the Romanian Academy, Romania, ORCID: 0009-0007-0647-8709, isabelle.oprea@gmail.com.

² Daniela Duță - Legal Research Institute "Acad. Andrei Rădulescu", School of Advanced Studies of the Romanian Academy (SCOSSAR); ORCID: 0000-0003-3335-9581, ghituleasad@yahoo.com.

1. Introduction

The digital transformation of the banking sector has reshaped how financial services are delivered, creating new opportunities to enhance operational efficiency and financial inclusion. The transition from traditional banking models to digital platforms has enabled faster and more accessible banking services.³

At the core of this transformation are artificial intelligence (AI)-powered virtual agents — such as chatbots and embodied conversational agents (ECAs) — which simulate human interaction using natural language processing (NLP), machine learning (ML), and data analytics. These technologies are increasingly integrated into customer service platforms, personal financial management tools, and sales processes, allowing banks to provide round-the-clock, scalable, and personalized assistance.⁴

Virtual agents — also referred to as AI assistants or AI agents — are intelligent software applications powered by artificial intelligence that simulate human-like interactions with users through text or voice-based communication. By leveraging AI and natural language processing, these agents can interpret user intent, deliver personalized responses, and improve over time through continuous learning. They are increasingly implemented in customer service environments to automate support, enhance self-service experiences, and assist human agents in delivering more efficient service.⁵

The use of virtual customer assistants as an alternative to customer service assistants is one of the most common use-cases of AI in banking.⁶

Financial inclusion refers to the ability of individuals and businesses to access a range of financial products and services such as bank accounts, savings, loans, and insurance that empower them to manage money effectively, including saving, receiving, and transferring funds.⁷

Despite growing global access, approximately 1.4 billion adults remained unbanked in 2021, most of them in developing economies.⁸ The COVID-19 pan-

³ Isabelle-Margareta Oprea, Liviu-Gelu Draghici (2024). „Bank Digitalization, Financial Literacy, and Inclusion in Romania”, *Manager*, 39(1): 22-38. The document is available online at: <https://manager.faa.ro/?p=8965>, accessed on 12.05.2025.

⁴ Hana Demma Wube, Sintayehu Zekarias Esubalew, Firesew Fayiso Weldesellase and Taye Girma Debelee (2022), „Text-based chatbot in financial sector: A systematic literature review”. *Data Science in Finance and Economics*, 2(3): 209–236. <https://doi.org/10.3934/DSFE.2022011>.

⁵ Jeanine Desirée Lund, *What is a virtual agent?*, 2025. The document is available online at: <https://www.puzzel.com/blog/what-is-a-virtual-agent#what-is-a-virtual-agent>, accessed on 12.05. 2025.

⁶ Daniela Duță, Isabelle Oprea, „The Role of Artificial Intelligence in the Digital Banking System”, in Cristina Elena Popa Tache, Renata Treneska Deskoska, Nathaniel Boyd (coordinating editors), *Adapting to Change Business Law insight from Today's International Legal Landscape*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2023, p. 230-244.

⁷ Isabelle-Margareta Oprea, Liviu-Gelu Draghici. *op. cit.*, p. 25.

⁸ World Bank, *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and*

demic further highlighted the importance of digital financial services in maintaining access during crises, particularly for underserved communities.⁹

AI-powered virtual agents are uniquely positioned to address financial exclusion by providing round-the-clock customer service, personalized financial education, and data-driven credit assessments that bypass traditional credit history requirements.¹⁰

For instance, virtual assistants such as chatbots and voice-based technologies are increasingly integrated into mobile banking applications to support users with financial transactions, expense management, and tailored product suggestions. Automated investment platforms, known as robo-advisors, allow individuals to handle investments and oversee their portfolios at significantly lower costs than traditional financial advisors, thereby expanding wealth-building access for users with limited incomes.

Beyond convenience, virtual agents also serve as a conduit for financial education. By simplifying complex banking terms and processes, these tools can reduce information asymmetry and empower users to make better financial decisions. However, the success of such implementations depends heavily on the technological infrastructure, user trust, and the bank's capacity to manage data securely and ethically.¹¹

Moreover, automation allows for more scalable, consistent, and efficient service delivery. Banks can personalize financial products using AI tools, based on real-time behavioral and transactional data, further offering customers the tools tailored to their needs.¹²

Despite the advantages, several risks accompany the growing reliance on virtual agents. One major concern is data privacy. As these agents collect and process sensitive customer information/data, ensuring data protection and privacy becomes critical. Improper handling of customer data can lead to breaches, ero-

Resilience in the Age of COVID 19, 2022, <https://doi.org/10.1596/978-1-4648-1897-4>. The document is available online at: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099818107072234182/idu06a834fe908933040670a6560f44e3f4d35b7>, accessed on 12.05.2025.

⁹ International Monetary Fund, *2023 Financial Access Survey*, 2023, The document is available online at: <https://data.imf.org/?sk=E5DCAB7E-A5CA-4892-A6EA-598B5463A34C>, accessed on 12.05.2025.

¹⁰ Consultative Group to Assist the Poor (CGAP), *AI's promise: A new era for financial inclusion*, 2023. The document is available online at: <https://www.cgap.org/blog/ais-promise-new-era-for-financial-inclusion>, accessed on 12.05.2025.

¹¹ Luay Anaya, Asma Braizat, Ria Al-Ani (2024), „Implementing AI-based Chatbot: Benefits and Challenges”. *Procedia Computer Science*, vol. 239: 1173-1179. DOI: 10.1016/j.procs.2024.06.284.

¹² Gang Kou, Pei Yang, Yi Peng, Feng Xiao, Yang Chen, Fawaz E. Alsaadi (2020), „Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods”. *Applied Soft Computing*, Volume 86, 105836, <https://doi.org/10.1016/j.asoc.2019.105836>.

sion of trust, and regulatory penalties. Studies highlight that privacy, data protection, cybersecurity threats, and unclear legal accountability are major deterrents to chatbot adoption in the financial sector.¹³

As virtual agents collect, store, and analyze sensitive financial information, the risk of data breaches and misuse increases, highlighting the need for strict adherence to regulations such as the General Data Protection Regulation¹⁴ and local data protection laws.¹⁵

The lack of transparency can erode trust, particularly among vulnerable groups who already face systemic barriers to accessing financial services. Ensuring fairness, accountability, and explainability in AI-driven decisions can prevent the reinforcement of existing biases and inequalities.¹⁶

Moreover, there are user experience challenges. While AI agents can handle routine inquiries effectively, their performance may deteriorate when faced with complex or emotionally nuanced interactions. Miscommunication, lack of empathy, and the “uncanny valley” effect where too-human avatars feel unsettling can negatively impact customer satisfaction.¹⁷

Financial institutions must therefore invest in inclusive user design and public education initiatives to ensure that technology serves as a bridge — not a barrier — to financial inclusion.

This paper aims to investigate how bank digitalization, particularly through the implementation of virtual agents, contributes to advancing financial inclusion while safeguarding data protection. To achieve this, the research is guided by two primary objectives: first, to explore how digital banking technologies, specifically virtual agents, facilitate financial inclusion; and second, to analyze the implications of these technologies on data protection. In alignment with these objectives, the study seeks to answer the following research questions: (1) How do virtual agents enhance financial inclusion? and (2) What are the risks and

¹³ Hana Demma Wube, Sintayehu Zekarias Esubalew, Firesew Fayiso Weldesellase and Taye Girma Debelee, *op. cit.*, p. 220.

¹⁴ 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.

¹⁵ Juan Carlos Crisanto, Cris Benson Leuterio, Jermy Prenio and Jeffery Yong, *Regulating AI in the financial sector: Recent developments and main challenges*, Bank for International Settlements, 2024. FSI Insights No. 63. The document is available online at: <https://www.bis.org/fsi/publ/insights63.htm>, accessed on 12.05.2025.

¹⁶ Reuben Binns (2018), „Fairness in machine learning: Lessons from political philosophy”. In Sorelle A. Friedler, Christo Wilson (eds.) *Proceedings of Machine Learning Research* 81:1–11, Conference on Fairness, Accountability, and Transparency, p. 149–159. The document is available online at: <https://proceedings.mlr.press/v81/binns18a.html>, accessed on 12.05.2025.

¹⁷ Maurice Wendt, Sebastian Schaefer, Mario Schaarschmidt (2025), *Empowering regional bank sales through embodied conversational agents: A multiple case study*, Spring 3-1-2025. Proceedings of SIGSVC Workshop 2024. The document is available online at: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1008&context=sprouts_proceedings_sigsvc_2024, accessed on 12.05.2025.

benefits related to data protection? These questions aim to frame the analysis and provide insight into the dual role those virtual agents play in modern banking systems, as both tools for inclusive financial outreach and potential vectors of data risk.

The research adopts a qualitative methodology through case studies of digital-first and traditional banks in Romania that have implemented AI-driven tools, analyzing both their successes and shortcomings. The research contributes to the growing discourse on digital finance by offering insights into how technology can shape a more inclusive, transparent, and secure financial ecosystem.

2. Case Study: Digital Banking and Virtual Agents in Romania

Romania has made notable strides in digital banking transformation, but the pace of adoption remains uneven across institutions. Technological solutions based on machine learning (ML) and artificial intelligence (AI) are primarily applied in areas such as data extraction and analysis, risk management, customer assessment through scoring models, and the detection and monitoring of fraud and anti-money laundering (AML) activities. Despite the fact that 81% of banks — accounting for 97% of the market share — view AI integration as a central element of their medium-term business strategies, the formal inclusion of these technologies within institutional governance structures remains limited.¹⁸

According to Chart 1, in Romania, AI is currently implemented by 9 banks, ML and cloud computing by 13 banks each, biometric systems for customer identification by 17 banks, optical character recognition (OCR) by 16 banks, while Big Data technologies are used by only 5 banks.¹⁹ This disparity highlights the uneven maturity in digital capabilities and underscores the necessity for strategic alignment between innovation, infrastructure, and regulatory readiness.

According to the most recent report of National Bank of Romania published in June 2024, the first three largest bank in Romania in terms of assets are Banca Transilvania, BCR and CEC Bank. Although the paper analyses these three banks regarding the implemented virtual agents there is to be noted that the first chat bot in Romania was launched on 13th of March 2018, by Libra Internet Bank, ranked on 13th place on the list of Romanian banks in terms of assets.²⁰

This pioneering virtual assistant was designed to facilitate user-friendly and fully online access to financial products, allowing individuals to open a bank

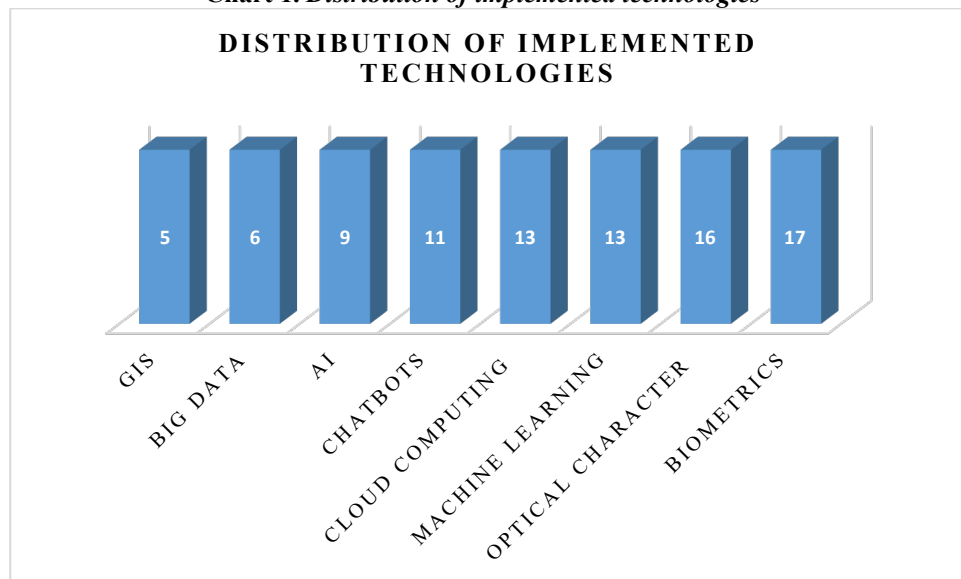
¹⁸ National Bank of Romania, *Raport asupra stabilitatii financiare*, 2024. The document is available online at: <https://bnro.ro/Raportul-asupra-stabilita%C8%9Bii-financiare---decembrie-2024-28873.aspx>, accessed on 12.05.2025.

¹⁹ Ibid.

²⁰ Ziarul Financiar, *BNR: Topul integral al băncilor din România după active în 2023. 15 bănci din 32 și-au schimbat poziția în top. CEC Bank a urcat cel mai mult, pe locul 3*, 2024. The document is available online at: <https://www.zf.ro/banci-si-asigurari/bnr-topul-integral-al-banci-lor-din-romania-dupa-active-in-2023-15-22412864>, accessed on 12.05.2025.

account, apply for loans, or obtain credit cards without visiting a physical branch. Within a short period, the chatbot recorded over 10,000 interactions with users interested in digital banking solutions.²¹

Chart 1. Distribution of implemented technologies



Source: representation of authors based on BNR data (2024)

2.1. Banca Transilvania

Banca Transilvania (BT), Romania's largest bank by assets, has been a national leader in implementing artificial intelligence (AI) and virtual agents within its operations. The bank's AI journey began in 2017 with the launch of its first chatbot for customer interaction, followed by rapid development and deployment of more advanced tools in subsequent years.²²

One of its first notable innovations was Livia, a virtual financial assistant integrated into the BT Pay app in 2018. Livia sends reminders about bills, offers

²¹ Libra Internet Bank, *Primul chatbot din România pentru vânzarea produselor bancare către clienți noi a avut peste 10 mii de interacțiuni în primele patru luni de la lansarea sa de către Libra Internet Bank*, 2018, March 13. The document is available online at: https://www.libra.bank.ro/Stiri/Primul_chatbot_din_Romania_pentru_vanzarea_produselor_bancare_catre_clienti_noi_a_avut_peste_10_mii_de_interactiuni_in_primele_patru_luni_de_la_lansarea_sa_de_catre_Libra_Internet_Bank/2110, accessed on 12.05.2025.

²² Banca Transilvania, *Chatbots of BT*, 2020. The document is available online at: https://www.banecatransilvania.ro/news-files/chatbots-of-bt-en/whitepapers_chatbotofbt_septembrie2020_en.pdf, accessed on 12.05.2025.

savings suggestions, and delivers real-time notifications, thereby increasing customer engagement and enabling financial self-management.²³

Accessible via Facebook Messenger, Skype, and phone, Livia has become a widely used tool, processing nearly 15 million messages and assisting over 90,000 customers. Its widespread adoption has contributed to a significant reduction in the volume of basic inquiries directed at human agents.²⁴

Building on this momentum, BT launched Raul in 2018, a chatbot specifically designed to support entrepreneurial clients. Raul assists with account information, outstanding balances, credit card limits, and other services essential to small business owners. Available across multiple platforms including Facebook Messenger, Skype, and WhatsApp, Raul has responded to approximately 830,000 messages and supported more than 20,000 entrepreneurs. The chatbot became especially valuable during the COVID-19 pandemic, enabling entrepreneurs to access key services remotely when physical interactions were restricted.²⁵

In response to the health crisis in 2020, the bank developed a new virtual assistant named Ino. Built in less than three days, Ino was launched to assist customers with questions about loan installment deferrals and the rescheduling process. Integrated into the BT website, Ino guided users through application steps and clarified terms of eligibility. Although designed as a temporary solution, Ino demonstrated the agility and scalability of BT's AI infrastructure, and plans were set in motion to expand its functionality to other banking areas such as NeoCont and customer data updates.²⁶

Internally, Banca Transilvania has also implemented AI solutions to support its workforce. David, an AI-based virtual assistant developed on DRUID's platform, automates tasks within the bank's operations helpdesk. Deployed across more than 500 BT offices, David performs tasks such as interest rate calculations, generating internal reports, and resolving IT issues. By reducing the time and errors associated with routine operations, David has improved employee productivity and service quality.²⁷

In the area of human resources, BT introduced Aida, a digital assistant that supports internal administrative processes. Aida helps employees manage documents, book appointments, and access internal HR services. As a task-oriented chatbot, Aida embodies BT's broader strategy of digital enablement across

²³ Ibid.

²⁴ Ibid.

²⁵ Ibid.

²⁶ BankingNews, *Banca Transilvania a lansat un nou chatbot. Ino este un asistent virtual dedicat clienților care vor informații despre amânarea ratelor*, 2020, April 30. The document is available online at: <https://bankingnews.ro/amanarea-platii-creditelor-banca-transilvania-chatbot-ino.html>, accessed on 12.05.2025.

²⁷ DRUID AI., *Banca Transilvania uses DRUID's AI agent to boost productivity*, 2023. The document is available online at: <https://www.druidai.com/case-studies/conversational-ai-chatbot-banking-employee-support-btrl>, accessed on 12.05.2025.

all operational layers.²⁸

In 2023, Banca Transilvania significantly expanded its AI capabilities by integrating Microsoft Azure OpenAI into its customer service platform, particularly on the “Întreb BT” (Ask BT) website. This platform uses AI-powered search functionality to deliver detailed answers to user questions, thus enhancing accessibility and improving the digital experience.²⁹

In 2024, BT further embedded AI into its internal workflows by adopting Microsoft 365 Copilot and GitHub Copilot. These AI assistants help improve productivity by automating document creation, summarizing data, and supporting software development teams.³⁰

Banca Transilvania’s commitment to enhancing digital banking experiences is further demonstrated by the integration of ChatBT, a conversational AI assistant embedded directly within the BT Pay mobile application. Launched in collaboration with DRUID in 2025, ChatBT offers real-time assistance to customers by answering a wide range of inquiries—from checking account balances and transaction history to providing personalized financial advice. The virtual agent also enables users to initiate payments, manage their cards, and receive guidance on banking procedures, all through a simple, user-friendly interface. One of the key innovations of ChatBT is its hybrid support functionality, which ensures seamless escalation to human agents when needed, maintaining both efficiency and personalization in customer interactions. The integration of ChatBT into BT Pay represents a strategic step toward human-centered automation, improving accessibility, convenience, and satisfaction for users who rely on mobile banking for their day-to-day financial management.³¹

As part of BT’s broader digital transformation agenda, ChatBT exemplifies the role of intelligent virtual agents in transforming mobile banking into a smart, responsive, and adaptive financial platform. Banca Transilvania is developing a unique feature in Romanian banking called “proactive support,” aimed at improving the efficiency of call center interactions. This system will enable operators to anticipate the reason behind a customer’s call even before the conversation begins. By analyzing real-time data and recent customer activities such as failed online payments, expired cards, or transaction errors, the bank can predict

²⁸ Banca Transilvania, *Chatbots of BT*. 2020. The document is available online at: White-Papers_Chats-Bot-of-BT_septembrie-2020_RO (bancatransilvania.ro), accessed on 12.05.2025.

²⁹ Banca Transilvania, *Banca Transilvania integrates Microsoft Azure OpenAI in customer communication*, 2023. The document is available online at: <https://en.bancatransilvania.ro/news/comunicare-de-presa/banca-transilvania-integreaza-microsoft-azure-openai-comunicarea-cu-clientii>, accessed on 12.05.2025.

³⁰ Banca Transilvania, *BT integrates Microsoft 365 Copilot and GitHub Copilot AI assistants*, 2024. The document is available online at: <https://en.bancatransilvania.ro/news/comunicare-de-presa/BT-integreaza-microsoft-365-copilot-si-github-copilot-asistenti-ai>, accessed on 12.05.2025.

³¹ DRUID AI, *DRUID AI Brings ChatBT to Banca Transilvania's BT Pay App*, 2025. The document is available online at: <https://www.druidai.com/news/druid-and-banca-transilvania-integrate-chat-bt>, accessed on 12.05.2025.

the issue and notify support staff of potential scenarios. The system will later be improved using artificial intelligence and machine learning to further refine its predictive capabilities.³²

2.2. Banca Comercială Română (BCR)

Banca Comercială Română (BCR) is the second largest bank in Romania in terms of assets and, in terms of digitalization, in the 2023 FinnoScore report, it achieved an impressive third place internationally, scoring 6.81 out of 10. This ranking reflects BCR's significant advancements in digital banking, particularly through its George mobile application, which has improved customer onboarding and online banking experiences.³³

This bank, has made substantial progress in integrating artificial intelligence (AI) into its operations, focusing primarily on customer engagement and operational efficiency. A key component of this digital transformation is the virtual assistant ADA, launched in 2022. ADA is an algorithmic artificial intelligence model based on generative AI technology, with its main function being to respond to user inquiries on the bank's website and the app, specifically related to BCR's Internet Banking and Mobile Banking services.³⁴

ADA, Banca Comercială Română's virtual assistant, offers comprehensive support for a wide range of banking needs, providing users with instant access to personalized information. It can display current account balances and IBANs, as well as details related to savings and term deposit accounts, including maturity dates and available funds. For loan management, ADA provides repayment schedules, outstanding balances, due dates, and options for early repayment or changing installment dates. It also supports credit card management, offering data on available balance, minimum payment, due dates, and plan subscriptions. Additionally, ADA can assist with card services such as blocking, PIN regeneration, limit adjustments, and delivery address changes. The assistant offers detailed updates on account garnishments and helps users register complaints or

³² Wall-Street, Future Banking Summit | *Banca Transilvania vrea să „citească gândurile” clienților care sună în call-center*, 2024. The document is available online at: <https://www.wall-street.ro/articol/Finante-Banci/311247/future-banking-summit-banca-transilvania-vrea-sa-citeasca-a-gandurile-clienților-care-suna-in-call-center.html>, accessed on 12.05.2025.

³³ Economedia, *Topul celor mai competitive bănci privind digitalizarea: BCR, Banca Transilvania și OTP Bank în top 3 internațional; BCR se află pe locul al treilea*, 2023, June 22. The document is available online at: <https://economedia.ro/topul-celor-mai-competitive-banci-privind-digitalizarea-bcr-banca-transilvania-si-otp-bank-in-top-3-international-bcr-se-afla-pe-locul-al-treilea.html>. Retrieved April 17, 2025, accessed on 12.05.2025.

³⁴ Banca Comercială Română, *BCR lansează chatbot-ul ADA, primul asistent virtual care poate oferi suport rapid și informații personalizate*, 2022. The document is available online at: <https://www.bcr.ro/ro/presa/informatii-de-presa/2022/04/14/BCR-lanseaza-chatbot-ul-ADA-primul-asistent-virtual-care-poate-oferi-suport-rapid-si-informatii-personalizate-atat-pentru-persoane-fizice-cat-si-solutii-de-finantare-pentru-companii>, accessed on 12.05.2025.

schedule branch visits, enhancing both convenience and digital self-service capabilities. It is accessible 24/7 via multiple platforms, including the BCR website, George Web, George Mobile, and WhatsApp, making it a versatile and easily accessible solution for clients.³⁵

ADA was designed to reduce pressure on human customer service representatives by providing fast, automated responses. According to BCR's 2024 half-year report, the virtual assistant handled over 455,500 conversations in the first six months of the year, with approximately 35% of those interactions fully automated, requiring no human intervention. ADA is capable of addressing nearly 2,000 types of inquiries, significantly enhancing self-service capabilities for customers.³⁶

In addition to ADA, BCR has implemented other AI-driven tools to further improve its customer service infrastructure. The Conversational Interactive Voice Response (IVR) system deployed in the BCR Contact Center delivered over 82,000 personalized responses in the first half of 2024 without requiring human assistance. This IVR system allows customers to speak naturally rather than navigating through traditional menu options, streamlining their experience and saving time.³⁷

Furthermore, BCR has introduced Voice ID, a biometric authentication feature based on voice recognition, which was used by nearly 266,000 customers during the same reporting period. This technology improves both security and convenience, enabling quick and secure customer identification during interactions with the Contact Center.³⁸

2.3. CEC Bank

CEC Bank, Romania's oldest bank, has actively pursued digital transformation in recent years. A key element of this strategy has been the implementation of virtual agents and low-code platforms to improve customer service and internal operations. CEC Bank introduced a virtual assistant named "Raluca" which was designed to improve customer service by offering real-time support for basic banking inquiries on the bank's official website. "Raluca" uses Google's Dialogflow platform for natural language processing, allowing it to interpret and respond in Romanian. Raluca's main function is to provide general information regarding CEC Bank's products and services. It can answer frequently asked questions, guide users through banking processes, and direct them to relevant

³⁵ Banca Comercială Română, *BCR Group H1 2024 results: Continuous support for raising financial literacy and multiplier investments in community*, 2024, August 2. The document is available online at: <https://www.bcr.ro/en/press/press-release/2024/08/02/BCR-group-H1-2024-results-continuous-support-for-raising-financial-literacy-and-multiplier-investments-in-community>, accessed on 12.05.2025.

³⁶ Ibid.

³⁷ Ibid.

³⁸ Ibid.

sections of the website. For example, it helps users understand how to open an account, locate ATMs or branches, and learn more about personal or business banking options. An important aspect of Raluca's design is its strict adherence to data privacy regulations. According to the bank's chatbot compliance statement, Raluca does not collect or store personal data. This ensures full alignment with the data protection regulation.³⁹

Users are informed clearly that any queries requiring personal banking data must be redirected to secure channels or physical branches. Raluca provides real-time support, significantly reducing the load on human customer service teams. Because the chatbot is accessible 24/7, customers can get answers to routine inquiries during weekends, holidays, or late-night hours. This enhances overall customer satisfaction and ensures that basic service remains uninterrupted, even outside of traditional banking hours. Currently, Raluca functions in a read-only informational capacity meaning it provides answers and guidance, but it does not perform direct transactions or account-specific actions. However, its successful implementation suggests that future versions might integrate more advanced AI capabilities, including secure identification and transaction processing, as long as these developments remain GDPR-compliant.⁴⁰

2.4. Comparative Analysis of Virtual Agents at Banca Transilvania, BCR, and CEC Bank

Virtual agents have become integral to the strategic evolution of Romania's banking sector, particularly among leading institutions like Banca Transilvania, Banca Comercială Română (BCR), and CEC Bank. While their functionalities often align with broader goals of digital efficiency, a closer look reveals distinct technological trajectories and strategic nuances.

A defining trend across all three banks is the operational convergence of AI-enabled automation with customer experience design. However, Banca Transilvania's architecture is notably more diversified. Its deployment of multiple agents tailored to specific business needs such as Raul for entrepreneurs and Aida for HR, illustrates a modular design strategy. This layered AI implementation allows for task-specific optimization, an approach consistent with global best practices in intelligent automation where organizations shift from single-agent deployment to orchestrated ecosystems.⁴¹

In contrast, BCR has positioned its virtual agent ADA not as an isolated chatbot, but as an extension of its George digital ecosystem. This integration re-

³⁹ CEC Bank, *Acord de conformitate – Chatbot – CEC Bank SA*, 2023, November 7. The document is available online at: <https://www.cec.ro/termeni-conditii/conditii-chat>, accessed on 12.05.2025.

⁴⁰ Ibid.

⁴¹ Hana Demma Wube, Sintayehu Zekarias Esubalew, Firesew Fayiso Weldesellase and Taye Girma Debelee, *op. cit.*, p. 221.

flects a holistic digital channel strategy where virtual agents act as customer-facing gateways to broader platform capabilities. From a digital transformation perspective, this aligns with research indicating that embedded AI within omnichannel systems improves both service cohesion and user retention.⁴²

ADA's implementation of biometric security and voice-based navigation further underscores BCR's investment in frictionless digital identity verification, a trend gaining prominence in European financial technology markets.⁴³

CEC Bank, on the other hand, adopts a minimalistic model characterized by compliance-first principles and limited functional scope. Its virtual agent, Raluca, operates as an informational layer without backend integration. While this limits transactional utility, it positions the bank favorably in terms of data protection and regulatory adherence — critical considerations amid increasing scrutiny over algorithmic governance in finance.⁴⁴

The differential pace and purpose of AI deployment in these banks also highlight a divergence in digital maturity. While large Romanian banks share similar goals of automation and customer-centricity, their internal investment capacity and technological infrastructure strongly influence outcomes. BT's AI roadmap, characterized by quick prototyping and user-responsive iteration, exemplifies agile governance frameworks rarely matched by institutions with more hierarchical or risk-averse cultures.⁴⁵

Furthermore, these virtual agents serve not only as tools for client interaction but as symbols of each bank's strategic orientation. Banca Transilvania's agents promote productivity and internal digitization, while BCR's ADA represents platform expansion and identity security. CEC Bank's Raluca, although limited in function, reflects a trust-centric model that prioritizes user safety over innovation.

3. Discussion

The integration of virtual agents in Romanian banking institutions — such as Banca Transilvania, BCR, and CEC Bank — demonstrates both the transformative potential and the complex challenges of AI-driven customer service platforms.

⁴² Oskar Bladh, Hedvig Henrekson & Ida Modée (2018). *The Impact of Virtual Agents on Customer Loyalty in Major Swedish Banks.*, 2018. The document is available online at: <http://www.diva-portal.org/smash/get/diva2:1213804/FULLTEXT01.pdf>, accessed on 12.05.2025.

⁴³ Luay Anaya, Asma Braizat, Ria Al-Ani, *op. cit.*, p. 1175.

⁴⁴ Zeyu Tang, Jiji Zhang, Kun Zhang (2023), „What-is and how-to for fairness in machine learning: A survey, reflection, and perspective.” *ACM Computing Surveys*, 2023, 55(13s): 1-37. <https://doi.org/10.48550/arXiv.2206.04101>.

⁴⁵ Daria Maria Sitea, Carolina Țîmbalari (2022), „Measuring the Banking Competitiveness. A Case Study of Romania”. *Revista Economică*, 74(2): 59-69, DOI: 10.56043/reveco-2022-0018.

3.1. Impact on Financial Inclusion

Virtual agents enhance financial inclusion in Romania by providing accessible, real-time, and user-friendly digital banking services that help overcome traditional barriers such as limited physical bank branches and low financial literacy.

By enabling 24/7 customer interactions, chatbots like BCR's ADA and BT's Livia offer autonomous banking assistance and perform banking operations independently at any time, such as checking balances, retrieving IBANs, managing security devices, and accessing product information, making banking more accessible especially for those who cannot easily visit branches.

In terms of cost reduction and scalability, virtual agents represent a major advancement. They reduce the need for large customer support teams and can scale seamlessly during peak times without performance loss and simplify the financial services for customers.

The deployment of virtual agents has had a tangible impact on financial inclusion in Romania. CEC Bank's Raluca chatbot, while limited in functionality, serves as an accessible entry point for first-time digital users, while BCR's ADA and BT's ecosystem enable deeper engagement through account management and advisory features. Digital financial services can reduce institutional biases and extend services to migrant workers, low-income groups, and rural populations.⁴⁶

Furthermore, these tools facilitate personalized financial education. AI algorithms integrated into virtual assistants are increasingly capable of offering tailored advice based on user behavior, spending patterns, or financial goals. This personalized approach is particularly important in addressing gaps in financial literacy — a persistent issue in Romania.⁴⁷

However, several challenges limit the full realization of these benefits. One primary issue is the digital literacy gap. Many users, particularly elderly populations and those in rural areas, may not possess the skills required to interact effectively with virtual banking platforms. Cybersecurity threats present another serious concern. As more sensitive financial transactions are conducted via virtual agents, the systems become prime targets for cyber-attacks.

⁴⁶ Claudiu Negrea, Ela Scarlat, „Challenges and opportunities for consolidating the availability of financial services in Romania”. In: *Economic growth in the conditions of globalization: conference proceedings: International Scientific-Practical Conference*, XVIth edition, October 12-13, 2022, Chisinau. Chisinau: INCE, 2022, volume II, pp. 159-173. ISBN 978-9975-3583-9-2; ISBN 978-9975-3385-6-1 (PDF). <https://doi.org/10.36004/nier.cecg.III.2022.16.13>.

⁴⁷ Isfandyar Zaman Khan, Natalie Nicolaou, Saniya Ansar, Juan Buchenau Hoth, Danilo Palermo Queiroz, Panayotis N. Varangis, (2020), *Financial Inclusion in Romania: Issues and Opportunities*. World Bank Group. Report. The document is available online at: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/830431587015032573/financial-inclusion-in-romania-issues-and-opportunities>, accessed on 12.05.2025.

3.2. Data Protection Concerns

Trust and user acceptance remain also very important. Despite improvements in natural language processing and chatbot design, many users still perceive human agents as more trustworthy and competent, especially for complex financial decisions. While chatbot service can approximate human interaction in some scenarios, many customers still prefer human agents, especially when trust and service complexity are at stake. The research showed that customer outcomes were influenced by levels of trust in chatbot service, and even light human intervention could improve satisfaction — implying a clear preference for human agents in contexts requiring confidence or emotional intelligence.⁴⁸

Additionally, when chatbot service fails — regardless of how “human-like” the chatbot appears — customers often switch to a human agent. This switching behavior is driven by expectations that only a human can manage recovery well in service failure scenarios. Thus, banks must invest in hybrid models that blend automated services with human support to foster trust and increase adoption.⁴⁹

In parallel, data protection challenges must also be considered, as virtual agents collect, store, and process large volumes of data. Privacy and data protection risks, such as unauthorized access, data breaches, or the misuse of personal information, can undermine user trust and may lead to legal sanctions under GDPR regulation.

To mitigate these risks, the financial institutions must implement strict cybersecurity measures, ensure transparency in data processing, and provide customers with real control over their personal information/ information, including the right to information, access, rectification, and erasure. In this context, digital trust becomes a fundamental pillar for the sustainable adoption of virtual agents in the banking sector.

3.3. Recommendations

The integration of virtual agents in Romanian banking, as explored through case studies of Banca Transilvania, BCR, and CEC Bank, highlights both transformative potential and systemic risks. The following recommendations target banks, policymakers, and developers to ensure that digital banking technologies support inclusive, secure, and ethical financial ecosystems.

⁴⁸ Rusty Stough, Dmitri Markovitch, Dongling Huang (2024), „Can chatbot customer service match human agents on customer satisfaction?” *Journal of Retailing and Consumer Services*. Volume 76, January, 1036002024, <https://doi.org/10.1016/j.jretconser.2023.103600>.

⁴⁹ Zhenzhen Lu, Qingfei Min, Lintong Jiang, Qi Chen (2024), „The effect of the anthropomorphic design of chatbots on customer switching intention when the chatbot service fails: An expectation perspective”. *International Journal of Information Management*, Volume 76, 102767, <https://doi.org/10.1016/j.ijinfomgt.2024.102767>.

For banks, one of the most significant barriers to effective use of virtual agents remains digital literacy, particularly among elderly populations and users in rural areas. Banks should invest in public-facing educational campaigns that provide users with the skills needed to confidently engage with AI-powered financial services. The success of chatbot integration depends not only on technological functionality but also on user readiness and training.⁵⁰

By offering tutorials, in-branch support for digital tools, and multilingual content, banks can increase adoption across all demographics. For long-term customer trust and institutional integrity there must be implemented ethical AI. Banks must proactively mitigate algorithmic bias, ensure transparency in decision-making processes, and embed fairness and accountability into AI systems. Fairness in AI is not merely a technical concern but an institutional imperative that demands cross-functional collaboration between IT, compliance, and legal teams.⁵¹ Romanian banks can lead in this space by publishing ethical AI guidelines, performing regular audits, and providing customers with clear information on how AI systems impact decisions related to loans, credit scoring, and service eligibility.

As virtual agents become increasingly embedded in banking operations, the cybersecurity threat surface also expands. Banks must strengthen their security protocols to protect sensitive customer data from breaches and misuse. Investments in encryption technologies, real-time threat detection, and multi-factor authentication are now industry standards. The Bank for International Settlements (2024) recommends a layered cybersecurity framework that includes AI-specific risk assessments, robust incident response systems, and employee training on cyber hygiene practices.⁵²

4. Conclusion

This study has examined the intersection of digital banking transformation and the implementation of virtual agents in Romania, with a focus on their dual role in fostering financial inclusion and protecting customer data. The research reveals both the opportunities and the limitations posed by AI-powered customer service solutions by analyzing the cases of Banca Transilvania, BCR, and CEC Bank.

Virtual agents have proven to be key enablers of financial inclusion by increasing accessibility, scalability, and personalization of financial services. Their ability to provide 24/7 assistance, deliver tailored financial education, and

⁵⁰ Hana Demma Wube, Sintayehu Zekarias Esubalew, Firesew Fayiso Weldesellase and Taye Girma Debelee, *op. cit.*, p. 223.

⁵¹ Zeyu Tang, Jiji Zhang, Kun Zhang, *op. cit.*, p. 20.

⁵² Bank for International Settlements, *Regulating AI in the financial sector: Recent developments and main challenges*, 2024, FSI Insights No. 63. The document is available online at: <https://www.bis.org/fsi/publ/insights63.htm>, accessed on 12.05.2025.

automate processes such as account inquiries and credit simulations has helped remove barriers that traditionally excluded rural, low-income, and underserved populations from formal banking systems. As demonstrated by tools such as BT's Livia and Raul, BCR's ADA, and CEC Bank's Raluca, virtual agents can effectively extend the banking network digitally — aligning with findings from CGAP and the World Bank, which highlight digital finance as a powerful tool for development.

Moreover, the study emphasizes that financial inclusion is no longer limited to mere access but extends to the quality and usefulness of the services provided. AI agents, when integrated into broader ecosystems — as exemplified by BCR's ADA within the George digital platform — can improve the autonomy of users and improve engagement through personalized, data-driven insights. However, the extent of these benefits remains contingent on digital literacy, infrastructural readiness, and institutional capacity to scale these tools equitably.⁵³

Simultaneously, the paper highlights the emerging risks that accompany the use of virtual agents — especially in terms of data protection. While automation introduces efficiency, it also amplifies vulnerability to data breaches and misuse. Institutions like CEC Bank have responded by limiting the data access capabilities of their virtual agents to ensure compliance with GDPR. In contrast, more advanced implementations, like BCR's voice biometric systems or BT's ChatBT, necessitate rigorous data governance frameworks to ensure transparency, ethical AI use, and customer trust.

Furthermore, the research addresses user trust and behavioral dynamics. Studies such as Huang et al.⁵⁴ and Lu et al.⁵⁵ confirm that despite advances in natural language processing and human-like interaction, users continue to prefer human agents — especially when dealing with emotionally complex or high-stakes transactions. This highlights the need for hybrid models where AI supports, rather than replaces, human interaction.

Ultimately, the paper contributes to both academic and practical understanding by presenting a comprehensive evaluation of how digital banking tools — when implemented responsibly — can contribute to a more inclusive and secure financial ecosystem. It also underscores the need for coordinated efforts among banks, regulators, and technology providers to build infrastructures that are both innovative and ethically sound.

As Romania's banking sector continues to digitalize, stakeholders must prioritize not only technological advancement but also user empowerment, regulatory alignment, and ethical considerations. The future of banking lies not just in automation, but in designing systems that promote equity, security, and resilience across all customer segments.

⁵³ Luay Anaya, Asma Braizat, Ria Al-Ani, *op. cit.*, p. 1175.

⁵⁴ Rusty Stough, Dmitri Markovitch, Dongling Huang (2024), *op. cit.*

⁵⁵ Zhenzhen Lu, Qingfei Min, Lintong Jiang, Qi Chen (2024), *op. cit.*

Bibliography

1. Anaya, Luay, Asma Braizat & Ria Al-Ani (2024), „Implementing AI-based Chatbot: Benefits and Challenges”. *Procedia Computer Science*, vol. 239: 1173-1179. DOI: 10.1016/j.procs.2024.06.284.
2. Binns, Reuben (2018), „Fairness in machine learning: Lessons from political philosophy”. In Friedler, Sorelle A. & Christo Wilson (eds.) *Proceedings of Machine Learning Research* 81:1–11, Conference on Fairness, Accountability, and Transparency, p. 149–159. The document is available online at: <https://proceedings.mlr.press/v81/binns18a.html>.
3. Bladh, Oskar, Hedvig Henrekson & Ida Modée (2018). *The Impact of Virtual Agents on Customer Loyalty in Major Swedish Banks.*, 2018. The document is available online at: <http://www.diva-portal.org/smash/get/diva2:1213804/FULLTEXT01.pdf>.
4. Crisanto, Juan Carlos, Cris Benson Leuterio, Jermy Prenio & Jeffery Yong, *Regulating AI in the financial sector: Recent developments and main challenges*, Bank for International Settlements, 2024. FSI Insights No. 63. The document is available online at: <https://www.bis.org/fsi/publ/insights63.htm>.
5. Duță, Daniela & Isabelle Oprea, „The Role of Artificial Intelligence in the Digital Banking System”, in Popa Tache, Cristina Elena, Renata Treneska Deskoska & Nathaniel Boyd (coordinating editors), *Adapting to Change Business Law in-sight from Today's International Legal Landscape*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2023, p. 230-244.
6. International Monetary Fund, *2023 Financial Access Survey*, 2023, The document is available online at: <https://data.imf.org/?sk=E5DCAB7E-A5CA-4892-A6EA-598B5463A34C>.
7. Khan, Isfandiyar Zaman, Natalie Nicolaou, Saniya Ansar, Juan Buchenau Hoth, Danilo Palermo Queiroz & Panayotis N. Varangis, (2020), *Financial Inclusion in Romania: Issues and Opportunities*. World Bank Group. Report. The document is available online at: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/830431587015032573/financial-inclusion-in-romania-issues-and-opportunities>.
8. Kou, Gang, Pei Yang, Yi Peng, Feng Xiao, Yang Chen & Fawaz E. Alsaadi (2020), „Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods”. *Applied Soft Computing*, Volume 86, 105836, <https://doi.org/10.1016/j.asoc.2019.105836>.
9. Lund, Jeanine Desirée, *What is a virtual agent?*, 2025. The document is available online at: <https://www.puzzel.com/blog/what-is-a-virtual-agent#what-is-a-virtual-agent>. Retrieved April 15, 2025.
10. National Bank of Romania, *Raport asupra stabilitatii financiare*, 2024. The document is available online at: <https://bnro.ro/Raportul-asupra-stabilitatii-financiare---decembrie-2024-28873.aspx>.
11. Negrea, Claudiu, Ela Scarlat, „Challenges and opportunities for consolidating the availability of financial services in Romania”. In: *Economic growth in the conditions of globalization: conference proceedings: International Scientific-Practical Conference*, XVIth edition, October 12-13, 2022, Chisinau. Chisinau: INCE, 2022, volume II, pp. 159-173. ISBN 978-9975-3385-6-1 (PDF). <https://>

- doi.org/10.36004/nier.cecg.III.2022.16.13.
12. Oprea, Isabelle-Margareta & Liviu-Gelu Draghici (2024). „Bank Digitalization, Financial Literacy, and Inclusion in Romania”, *Manager*, 39(1): 22-38. The document is available online at: <https://manager.faa.ro/?p=8965>.
13. 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.
14. Sitea, Daria Maria & Carolina Țîmbalari (2022), „Measuring the Banking Competitiveness. A Case Study of Romania”. *Revista Economică*, 74(2): 59-69, DOI: 10.56043/reveco-2022-0018.
15. Stough, Rusty, Dmitri Markovitch & Dongling Huang (2024), „Can chatbot customer service match human agents on customer satisfaction?” *Journal of Retailing and Consumer Services*. Volume 76, January, 1036002024, <https://doi.org/10.1016/j.jretconser.2023.103600>.
16. Tang, Zeyu, Jiji Zhang & Kun Zhang (2023), „What-is and how-to for fairness in machine learning: A survey, reflection, and perspective.” *ACM Computing Surveys*, 2023, 55(13s): 1-37. <https://doi.org/10.48550/arXiv.2206.04101>.
17. Wendt, Maurice, Sebastian Schaefer & Mario Schaarschmidt (2025), *Empowering regional bank sales through embodied conversational agents: A multiple case study*, Spring 3-1-2025. Proceedings of SIGSVC Workshop 2024. The document is available online at: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1008&context=sprouts_proceedings_sigsvc_2024.
18. World Bank, *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID 19*, 2022, <https://doi.org/10.1596/978-1-4648-1897-4>.
19. Wube, Hana Demma, Sintayehu Zekarias Esubalew, Firesew Fayiso Weldesellase and Taye Girma Debelee (2022), „Text-based chatbot in financial sector: A systematic literature review”. *Data Science in Finance and Economics*, 2(3): 209–236. <https://doi.org/10.3934/DSFE.2022011>.
20. Zhenzhen Lu, Qingfei Min, Lintong Jiang, Qi Chen (2024), „The effect of the anthropomorphic design of chatbots on customer switching intention when the chatbot service fails: An expectation perspective”. *International Journal of Information Management*, Volume 76, 102767, <https://doi.org/10.1016/j.ijinfo mgt.2024.102767>.

REGULATION, DATA PROTECTION AND AI GOVERNANCE IN THE EU

The New EU Product Liability Directive. Interaction with Parallel EU Initiatives: Proposed AI Liability Directive, Digital Services Act and Digital Markets Act

Assistant professor **Dimitrios DEVETZIS**¹

Abstract

On 18 November 2024, the European Union adopted the new Product Liability Directive – Directive (EU) 2024/2853 – to replace its nearly 40-year-old predecessor (85/374/EEC). This overhaul was driven by the need to update strict liability rules for products in light of digital technologies, artificial intelligence (AI), and new supply chain models. The old regime from 1985 had become “ill-suited to the digital age,” leading to gaps and legal uncertainty (for example, whether standalone software is a “product” under the law). The new Directive aims to ensure that injured persons enjoy the same level of protection irrespective of the technology involved, while businesses benefit from clearer rules and a level playing field. It introduces significant changes: expanding the definition of “product” to include digital and intangible items, broadening the range of liable persons beyond traditional manufacturers, and easing the burden of proof for claimants in complex cases. This modernized framework not only strengthens consumer protection in the internal market but also seeks to maintain fairness by balancing innovation incentives with accountability for harm. The essay that follows provides an overview of the key provisions of the new Product Liability Directive (“PLD”), analyzes its legal and doctrinal innovations, and examines its interplay with parallel EU initiatives such as the proposed AI Liability Directive, the Digital Services Act (DSA) and Digital Markets Act (DMA). Detailed footnotes and a consolidated bibliography are included to support the analysis of this important development in European product liability law.

Keywords: EU Product Liability Directive, AI Liability Directive, Digital Services Act (DSA), Digital Markets Act.

JEL Classification: K15, K22, K24

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/07>

Please cite this article as:

Devetzis, Dimitrios, „The New EU Product Liability Directive. Interaction with Parallel EU Initiatives: Proposed AI Liability Directive, Digital Services Act and Digital Markets Act”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 111-143.

¹ Dimitrios Devetzis - Frederick University of Cyprus; Visiting Professor at Université Orthodoxe au Congo; Legal Specialist on Law & Technology and civil Law, European Commission, ORCID: <https://orcid.org/0009-0003-1811-0451>, dimdeve.dd@gmail.com.

1. Introduction

After nearly forty years under the 1985 Product Liability Directive (Directive 85/374/EEC), the European Union has enacted a sweeping reform of its product liability framework. The original directive introduced a regime of *strict liability* for defective products, meaning a producer could be held liable for damage caused by a defect in its product irrespective of fault. While path-breaking for its time, the 1985 framework increasingly struggled to keep pace with technological and market developments in subsequent decades. Questions arose as to whether intangible digital products like software were covered, how to assign responsibility in complex global supply chains, and how injured persons could overcome evidentiary hurdles posed by emerging technologies (such as AI “black box” systems). The need for reform was broadly recognized by EU institutions and commentators alike. In 2022, the European Commission proposed a new Product Liability Directive to “modernise liability rules for the digital age” and address gaps in victim protection and legal clarity². Following negotiations, Directive (EU) 2024/2853 on liability for defective products was adopted on 23 October 2024, published in the Official Journal on 18 November 2024, and entered into force in December 2024³. EU Member States now have until 9 December 2026 to transpose the new rules into national law⁴.

The new PLD repeals the 1985 Directive and retains the core principle of *no-fault liability* of producers for defective products causing damage. However, it significantly updates and expands the scope of that regime. The reform reflects several policy objectives: ensuring the rules are fit for the digital age (covering software, AI and other digital products), fit for global value chains (ensuring an EU-domiciled defendant is available to consumers even when products come from abroad), and providing better protection for victims with le-

² Michael G. Faure, ‘Product Liability and Product Safety in Europe: Harmonization or Differentiation?’ *Kyklos*, 53 (2000) 4: 467–508, pp.498–499; Jacquemin, Z., ‘Product Liability Directive: Disclosure of Evidence, the Burden of Proof and Presumptions’, *Journal of European Tort Law*, 2024, 126–139; Koch, B. A., ‘Product Liability on the Way to the Digital Age’, *Journal of European Tort Law*, 2024, 109–125.

³ Piovano, Ch., Hess Ch., *Das neue europäische Produkthaftungsrecht – EU-Produkthaftungsrichtlinie (ProdHaftRL)*. Nomos Verlagsgesellschaft, 2024, p. 25; See also Tahoori Heydari, ‘A Review of the Product Liability Directive and the Proposal for a Directive of Liability for Defective Products’, (2024) *Al-Zaytoonah University of Jordan Journal for Legal Studies*, Special Issue, 962–971, 968–969; Rohrißen B., *Die EU-Produkthaftungs-RL 2024: Der „final compromise text“ Verschärfte Produkthaftung plus Product Compliance-Pflichten im Zeichen von Digitalisierung, KI und Globalisierung*, accessible on: https://www.nomos.de/wp-content/uploads/2024/02/NL-Product-Compliance_Februar-24_Zeitschriften-Archiv_Rohrssen_GesamtPDF.pdf#:~:text=mehrfach%20verschärft,dem%20Parlament%20übersandt%2C%20nachdem (last access (02.05.2025)).

⁴ Gitta Veldt, ‘The New Product Liability Proposal – Fit for the Digital Age or in Need of Shaping up?’, *EuCML* 1 2023, 24–31, 26.

gal certainty for industry (through tools like evidence disclosure and presumptions to alleviate the burden of proof)⁵. In parallel, the EU has been advancing other legal frameworks relevant to digital markets and AI. Notably, the Digital Services Act (Regulation (EU) 2022/2065) and Digital Markets Act (Regulation (EU) 2022/1925) came into force in 2022, reshaping platform obligations and competition in the digital economy. Moreover, the EU adopted the AI Act (Regulation (EU) 2024/1689) – a comprehensive product safety regulation for AI systems – and considered a complementary AI Liability Directive for fault-based liability in AI contexts⁶. The new PLD forms part of this broader landscape of EU digital regulation, addressing the civil liability dimension of harms caused by products in a technology-driven world.

This essay proceeds to analyze the new Product Liability Directive in a structured manner. First, it highlights the major legal and doctrinal innovations introduced, contrasting them with the prior regime. It then examines specific substantive changes, including the expanded definition of “product” and the enlarged circle of liable actors (such as online platforms and fulfilment service providers)⁷. Next, it discusses the Directive’s new provisions on burden of proof and disclosure of evidence, which aim to rebalance informational asymmetries in complex cases. The analysis will also consider the interplay between the PLD and the proposed AI Liability Directive, as well as its relationship with the DSA and DMA – situating the Directive within the EU’s wider digital governance framework. Finally, the essay addresses challenges and considerations for implementation, including potential uncertainties, the Directive’s impact on innovation and litigation, and the steps needed for a smooth transposition into national laws. The goal is to provide legal scholars with a cohesive, in-depth understanding of Directive 2024/2853 and its significance for European product liability doctrine and practice.

2. Legal and Doctrinal Innovations

Directive 2024/2853 introduces a number of legal and doctrinal innovations that mark a new era in EU product liability law. At its core, the Directive preserves the fundamental principle of strict liability for producers of defective products – a hallmark of the 1985 regime – but it modernises the scope, definitions, and procedural mechanisms around this principle. One key innovation is the Directive’s full harmonisation approach. Whereas the 1985 Product Liability

⁵ Masnada M., Pacciti A., Ecanova C., *EU introduces comprehensive digital-era Product Liability Directive*, 2024, accessible on: <https://www.hoganlovells.com/en/publications/eu-introduces-comprehensive-digitalera-product-liability-directive#:~:text=when%20online%20platforms%20function%20solely,once%20again%2C%20consistently%20with%20the> (last access: 02.05.2025).

⁶ See also Tahoora Heydari, *op. cit.*, p. 968-969.

⁷ See also Gitta Veldt, *op. cit.*, p. 26; Piovano, Ch., Hess Ch., *op. cit.*, p. 30.

Directive allowed certain Member State derogations (most notably on the “development risks” or state-of-the-art defense)⁸, the new PLD seeks maximum harmonization by preventing Member States from enacting national provisions that are either more or less stringent than the Directive’s rules⁹. The Directive thus aims to create a uniform standard of liability across the EU, enhancing legal certainty and fairness in cross-border cases.

The new PLD also updates the *doctrinal understanding of what constitutes a product defect and who qualifies as a producer*. It explicitly addresses technological developments by recognizing that products may have digital components and software that evolve after being placed on the market. For example, Article 6 of the Directive expands the criteria for defectiveness: when assessing whether a product is defective (i.e. does not meet the safety one is entitled to expect), courts must now take into account factors such as the product’s ability to learn or adapt after deployment (as in AI systems), the influence of other products (including software or connectivity) on the product, and compliance with any relevant cybersecurity or safety requirements¹⁰. This means that the concept of “defect” is no longer frozen at the moment of sale¹¹; if a product later acquires new functions or risks (say, via a machine-learning update), those can be considered in determining defectiveness. The relevant point in time for judging defectiveness may extend to when a product was *put into service* or when it *left the manufacturer’s control*, rather than only when it was first placed on the market¹². Doctrinally, this adapts the strict liability model to products whose characteristics change over time, ensuring the law remains effective for technologies like AI and IoT (Internet of Things) devices.

Another significant innovation is the introduction of procedural mechanisms to better balance the interests of injured persons and defendants. The Directive builds on the observation that information asymmetry and scientific complexity have made it excessively difficult for consumers to prove defects and causation under the old regime. In response, the PLD incorporates two novel tools common in some national systems but unprecedented at the EU level for product liability: court-ordered disclosure of evidence and rebuttable presumptions that ease the burden of proof for claimants. These mechanisms, discussed in detail in

⁸ Gitta Veldt, *op. cit.*, p. 26; J. Triaille ‘The EEC Directive of July 25, 1985 on Liability for Defective Products and Its Application to Computer Programs’ (1993) Computer Law and Security Report, 215, at 215 and 220; Lawrence C. Mann and Peter R. Rodrigues, ‘The European Directive on Products Liability: The Promise of Progress?’ (1988) 18 391 *Georgia Journal of International and Comparative Law*, 391-426, pp. 404-405.

⁹ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

¹⁰ V. Burgsdorff Christoph, *Increased liability due to the new EU Product Liability Directive: what does this mean for the medical and pharmaceutical industry?*, accessible on: <https://www.ibanet.org/increased-liability-eu-product-directive> (last access: 02.-5.2025); Piovano, Ch., Hess Ch., *op. cit.*, p. 40.

¹¹ See also for the old ‘status-quo’: Lawrence C. Mann, Peter R. Rodrigues, *op. cit.*, p. 404-405.

¹² Masnada M., Pacciti A., Ecanova C., *op. cit.*

a later section, represent a shift in doctrine towards facilitating *private enforcement* of product liability claims¹³. They reflect a policy choice to tilt the scales somewhat in favor of consumers (claimants) in recognition of the complexity of modern products. This is a notable evolution from the 1985 Directive, which left burdens of proof entirely to national rules and did not contemplate such evidentiary aids.

The new Directive also doctrinally extends the *protective purpose* of product liability law to cover new forms of harm. Notably, it expressly includes damage to data as a compensable harm caused by defective products. Under the 1985 regime, recoverable damage was basically limited to death, personal injury, and property damage (with certain exclusions and thresholds)¹⁴. The 2024 Directive adds “destruction or corruption of data” as a category of material damage for which compensation can be sought¹⁵. This acknowledges that in the digital era, a defective product (e.g. a malfunctioning software update or IoT device) might wipe out or compromise important data, causing real economic loss. The Directive makes clear that such data loss is to be treated akin to property damage (with the caveat that if the data can be restored or recovered at no cost, or if it is used for professional purposes, compensation under the Directive may be limited)¹⁶. Doctrinally, this inclusion of data damage signals an expansion of the traditional scope of product liability to intangible interests, aligning the law with the realities of modern consumer harm.

Finally, the PLD demonstrates a re-calibration of the balance between innovation and accountability. EU lawmakers have emphasized that while the law must protect consumers, it should also “give legal clarity and a level playing field to producers” and not unduly stifle innovation¹⁷. One area where this balance is evident is the treatment of the “development risks” defense (also known as the state-of-the-art defense). The 1985 Directive allowed Member States to decide whether or not to permit a producer to escape liability if the scientific and technical knowledge at the time was insufficient to discover the defect (thus, some countries like France disallowed this defense to favor consumers, while others allowed it)¹⁸. The new Directive moves to harmonize this: it generally allows the development-risk defense across all Member States, meaning a producer is *not liable* if it proves that the objective state of scientific and technical knowledge when the product was under its control was not such that the defect could be discovered¹⁹. This EU-wide acceptance of the defense has raised concern among some commentators that consumers in jurisdictions which previously barred this

¹³ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

¹⁴ J. Triaille *op. cit.*, p. 215 and 220.

¹⁵ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

¹⁶ Gitta Veldt, *op. cit.*, p. 26.

¹⁷ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125; Rohrießen B., *op. cit.*

¹⁸ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

¹⁹ V. Burgsdorff Christoph, *op. cit.*; Rohrießen B., *op. cit.*

defense may lose a layer of protection. However, the Directive also carves out important exceptions: a producer cannot invoke the development-risk defense if the defectiveness of the product is due to a related service, software (including updates), lack of required software updates, or a substantial modification of the product²⁰. In other words, for certain modern scenarios – such as unsafe software or failure to update AI systems – the producer *will* be held liable even if the risk was not discoverable at the time, effectively prioritizing consumer safety in those contexts. This nuanced approach illustrates the Directive’s doctrinal attempt to balance fostering innovation (by not imposing liability for truly unknown risks in core areas) against ensuring accountability when foreseeable digital-related risks are at play.

In sum, the legal innovations of Directive 2024/2853 can be seen as a comprehensive update of the EU’s strict product liability model. They preserve the fundamental risk-allocation logic (the producer is in principle the best bearer of the risk of product defects), while updating definitions, scope of damages, and procedural rules to address the challenges of new technologies and complex supply chains. The next sections delve into several of these changes in greater detail, beginning with the expanded notion of what constitutes a “product” under the Directive.

3. Expanded Definition of Product

One of the cornerstone changes in the new Product Liability Directive is the expanded definition of “product.” The 1985 Directive defined “product” in fairly narrow, tangible terms – essentially as movable goods (and electricity)²¹. This left ambiguity over whether intangible items like software could be considered products, especially when supplied independently of any physical medium. Divergent interpretations emerged across Member States over the years as software and digital content became ubiquitous, leading to legal uncertainty. Directive 2024/2853 addresses this gap decisively by broadening the definition to include digital and intangible elements. According to Article 4(1) of the new Directive: “*‘Product’ means all movables, even if integrated into or inter-connected with another movable or an immovable; it includes electricity, digital manufacturing files, raw materials and software.*”²². This clarifies that software is expressly included as a product for liability purposes, whether it is embedded in a physical good or provided as a standalone. “Digital manufacturing files” (for example, CAD files or design files used in 3D printing) are also included – these

²⁰ Masnada M., Pacciti A., Ecanova C., *op. cit.*; Rohrießen B., *op. cit.*

²¹ Shu Li and Beatrice Schutte, ‘The Proposal for a Revised Product Liability Directive: The Emperor’s New Clothes? (2023) *Maastricht Journal of European and Comparative Law* 30 (5): 573–596, 592.

²² Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

are essentially data files that can be used to produce a tangible item, and the Directive treats them as products because a defect in such a file could yield a defective physical product²³.

By expanding “product” in this manner, the Directive ensures that the regime of strict liability extends to the kinds of digital products and software that play a central role in today’s economy. For instance, a standalone software application that causes damage (say, a medical diagnosis app giving dangerously wrong advice, or malware-like behavior causing data loss) can now trigger producer liability just as a defective physical device could. This was a deliberate response to earlier uncertainty; the Commission and legal experts had noted that excluding software from the product definition was a significant shortcoming of the old framework²⁴. The inclusion of AI systems follows as a consequence – AI software is now a product, and if an AI system (e.g. an autonomous driving algorithm) is defective and causes damage, the PLD applies. The Directive’s recitals emphasize that even software provided in exchange for personal data (rather than a monetary price) should be covered, as long as it is provided in the course of a commercial activity²⁵. This means that the increasingly common business model of “free” digital services paid for with user data does not escape product liability if the software is defective; it closes a loophole where a company might argue no “product sale” occurred.

Importantly, the Directive also speaks to products that have related digital services and embedded software. Recital provisions clarify that integrated or interconnected digital services that are essential for a product’s functioning and safety are within scope²⁶. For example, consider a smart thermostat that relies on a cloud service to regulate temperature: that cloud service is an integral part of the product’s overall safety. If a defect in the service (say, a server outage or error) causes damage (e.g. pipes burst from lack of heat), it would be treated as part of the product’s defect. Similarly, the Directive ensures that open-source software components, when integrated into commercial products, do not fall through the cracks of liability – the commercial entity that integrated the open-source component can be treated as a producer of the overall product²⁷. The broad message is that the form of the technology (tangible or intangible) is irrelevant to the application of strict liability; what matters is the product’s role in causing damage.

Another aspect of the expanded scope is recognizing components and

²³ Gitta Veldt, *op. cit.*, p. 26.

²⁴ Masnada M., Pacciti A., Ecanova C., *op. cit.*; Wachter S., „Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the EU, the US, and Beyond”, *Yale Journal of Law & Technology*, vol. 26, no. 3, 2024, accessible on: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4924553 (last access: 02.05. 2025).

²⁵ Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

²⁶ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125; Wachter S., *op. cit.*

²⁷ Shu Li, Beatrice Schutte, *op. cit.* p. 592.

raw materials more explicitly. While the old directive implicitly covered components (by making their manufacturer liable as a producer if the component was defective), the new PLD is clearer in covering “components of products” in the definition of product²⁸. This dovetails with provisions on liable actors (discussed below) which explicitly include component manufacturers. Likewise, raw materials – say chemicals or ingredients – are listed as products, ensuring that if a raw material is defective and causes harm when incorporated into a final product, the raw material supplier could be held liable.

Additionally, as part of being “fit for the circular economy,” the Directive considers cases of products that are remanufactured or significantly modified after initial production. While this is addressed more directly in the context of who is liable (the person who substantially modifies a product can become the new ‘producer’ of that modified product), it also ties into the definition of product in a temporal sense. A product may undergo upgrades or refurbishments; the law needs to capture its state post-modification. The Directive doesn’t change the definition of product per se to include services or modifications, but by extending liability to modifiers, it effectively treats the modified item as a new “product” under the responsibility of the modifier²⁹. The definition of product is also tied to the notion of when a product is “put into circulation.” The old directive’s use of that term led to ambiguities (e.g. is a product in circulation when it leaves the factory, when it’s sold to a distributor, or to a consumer?). The new Directive uses updated terminology – “placed on the market or put into service” – and even notes that if a manufacturer retains control (such as the ability to update or withdraw a product remotely), the relevant time for defect assessment can be when it left the manufacturer’s control³⁰. While not exactly part of the *definition* of product, this temporal extension reinforces that products are seen as potentially dynamic and subject to post-sale changes.

In conclusion, the expanded definition of “product” under Directive 2024/2853 ensures that the EU product liability regime comprehensively covers the modern spectrum of goods and digital products. By explicitly including software, digital files, and related services, the law closes a long-recognized gap and brings intangible products into parity with traditional goods in terms of liability exposure. This change aligns with parallel legal developments (such as the inclusion of software as a “product” under the EU’s emerging AI and digital regulations) and ensures that consumers are not left unprotected simply because a defective “product” came in the form of code or data rather than plastic or metal. It also gives manufacturers and developers clear notice that quality and safety duties extend to their digital offerings. Having considered what products are covered, we turn next to *who* can be held liable – as the Directive also broadens the range of liable actors in response to global and platform-based supply chains.

²⁸ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

²⁹ Masnada M., Pacciti A., Ecanova C., *op. cit.*

³⁰ Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

4. Liable Actors, Including Platforms and Fulfillment Services

The new Product Liability Directive expands the circle of **liable actors** (or “economic operators” in product distribution) to ensure that an injured person can always find at least one solvent defendant within the EU from whom to claim compensation³¹. Under the 1985 Directive, the primary liable party was the *producer* of the defective product, defined to include the manufacturer of a finished product, the producer of any raw material or component, and anyone who presents themselves as producer by putting their name or brand on the product (the so-called “own-brand” or quasi-producer)³². Importers of products into the EU were also held liable if the manufacturer was outside the EU. While that scheme worked for many scenarios, the globalization of supply chains and rise of e-commerce left certain liability gaps – for instance, products directly shipped from non-EU producers to consumers, or scenarios involving online marketplaces and fulfillment centers that were not clearly covered by the old definitions. Directive 2024/2853 addresses these issues by introducing a graduated system of liable persons in Article 7, which can be summarized as follows:

- **Manufacturers and Quasi-Manufacturers:** The manufacturer of the defective product remains primarily liable, as before. This includes not only the maker of the final product but also the manufacturer of a defective component or ingredient that is integrated into a final product³³. If a company puts its trademark or name on a product made by someone else, it is treated as a manufacturer (quasi-producer) and is liable. A significant addition is that any person who substantially modifies a product after it has been placed on the market or put into service is now considered a manufacturer of the modified product³⁴. This targets “circular economy” scenarios: for example, a company that refurbishes or upgrades a used machine (outside the original manufacturer’s control) and in doing so introduces a defect can be held liable as a producer of the new version. This incentivizes those who remanufacture or heavily alter products to ensure safety, since they cannot simply point back to the original producer.

³¹ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125; Meryll Hervieu, *Point sur la nouvelle directive européenne (UE) 2024/2853 relative à la responsabilité du fait des produits défectueux*, Dalloz, 2025, accessible on: https://actu.dalloz-etudiant.fr/a-la-une/article/point-sur-la-nouvelle-directive-europeenne-ue-20242853-relative-a-la-responsabilite-du-fait-de/h/256e035c15335593d9c1bb38f7809c83.html?utm_source=chatgpt.com (last access: 02.05.2025); Sylvie Gallage-Alwis and Gaetan de Robillard, *Product regulation and liability in France* (Signature Litigation, 2024) accessible on: <https://www.signaturlitigation.com/sylvie-gallage-alwis-and-gae-tan-de-robillard-discuss-product-regulation-and-liability-in-france-in-lexology> (last access: 02.05.2025).

³² V. Burgsdorff Christoph, *op. cit.*

³³ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125. For the previous state of play compare the remarks of Michael G. Faure, *op. cit.*, pp. 498–499.

³⁴ Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

- **Importers and Authorized Representatives:** If the manufacturer is outside the EU, the **importer** of the product into the EU is liable, as under the old regime. The new Directive extends equal liability to the manufacturer's *authorized representative* in the EU³⁵. An authorized representative is defined as any person or entity established in the EU with a written mandate from the manufacturer to act on the manufacturer's behalf in relation to specified tasks (often this concept appears in product safety regulations). Now, such representatives – for example, an EU-based agent responsible for EU compliance – can be directly sued if the product they represent is defective. This is logical given that many modern regulations (like the EU Machinery Regulation or Medical Devices Regulation) already require non-EU manufacturers to designate an EU representative. The PLD leverages that by making the representative share liability, which ensures non-EU companies cannot evade liability by hiding abroad.

- **Fulfilment Service Providers:** A major innovation is the inclusion of *fulfilment service providers* as potentially liable parties, on a subsidiary basis. Article 7 provides that if neither the manufacturer, nor an importer, nor an authorized representative is present in the EU, then a fulfilment service provider (FSP) involved can be held liable³⁶. Fulfilment service providers are defined as entities offering at least two of the following services: warehousing, packaging, addressing, and dispatching of products, without having ownership of the products³⁷. This category is clearly aimed at scenarios like Amazon's "Fulfillment by Amazon" or third-party logistics companies that handle distribution for foreign sellers. Under the old law, if a consumer bought a product from outside the EU via an online marketplace, and there was no EU importer (the consumer effectively imported it themselves), they might find no one to sue in the EU if the product was defective. Now, the company that facilitated getting that product to the consumer – the FSP – can be on the hook if the upstream producer/importer cannot be identified in the EU. The policy rationale, as noted in commentary, is to avoid a liability vacuum and to compel those who profit from enabling market access to shoulder responsibility when needed³⁸. However, observers have noted a potential concern: some fulfilment providers are essentially logistics firms with no technical knowledge of the products they handle. Imposing liability on them might seem harsh since they cannot easily assess product safety. The counterargument is that this pressure will encourage fulfilment providers to deal only with suppliers who have EU-based responsible entities or to insure against such risks.

³⁵ See also 7 (2) of Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products, COM (2022) 495 FINAL.

³⁶ See European Parliamentary Research Service, 'Revised Product Liability Directive' (Briefing, European Parliament, February 2025) accessible on: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)739341](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)739341) (last access: 02.05.2025).

³⁷ Sven Förster, Dardan Gashi (2024), *The EU's new Product Liability Directive (from a German perspective)*, <https://www.clydeco.com/en/insights/2024/04/the-eu-s-new-product-liability-directive>.

³⁸ Michael G. Faure, *op. cit.*, p. 498-499.

It also complements the Market Surveillance Regulation (EU) 2019/1020, which already requires an “economic operator” in the EU for certain imports and names fulfilment service providers as a fallback responsible party for compliance issues³⁹.

- **Distributors and Online Platform Providers:** Finally, the Directive allows that distributors of the product and online platforms can be held liable *under certain conditions* as a last resort⁴⁰. This applies when no manufacturer, importer, authorized representative, or fulfilment provider can be identified in the EU. In such a case, a victim can potentially turn to the distributor who sold the product or the online platform that facilitated the sale, provided certain additional conditions are met. The Directive’s text and recitals indicate that online platforms are liable only when they *effectively act as an economic operator* in the chain, rather than a neutral intermediary. If an online marketplace presents itself as the seller, or otherwise does more than just enable a third-party sale (for example, if it handles fulfillment and marketing to a degree that it “assumes the role” of a distributor), it can be treated as such for liability purposes⁴¹. Conversely, if the online platform truly acts only as an intermediary – a passive hosting service for others’ listings – then its liability is governed not by the PLD but by the Digital Services Act (which preserves the e-commerce safe harbor for mere intermediaries)⁴². In essence, the PLD says to online platforms: if you behave like a seller or distributor, you will be treated as one; if you are genuinely just a broker, you won’t be liable under this Directive (though you must still abide by the DSA’s requirements, such as vetting sellers). This nuanced approach prevents undermining the DSA’s intermediary liability protections⁴³, while also preventing platforms from escaping liability when they are in practice deeply involved in the transaction. A practical example might be instructive: If a consumer buys a gadget from an online marketplace and that gadget is defective and causes harm, and it turns out the manufacturer is outside the EU and no importer or rep is present, then if the marketplace stored and shipped the item (fulfilment role) or presented it under its own branding, the consumer could sue the marketplace under the PLD. If the marketplace only connected the buyer and seller and the item was sent directly by a third-party seller, the marketplace might invoke the DSA safe harbor – though the consumer might then look to any fulfilment provider or ultimately face the difficulty of having no EU defendant (which is exactly the scenario the Directive tries to minimize). The DSA also imposes a “Know Your Business Customer” obligation on online marketplaces to obtain seller information and mandates a system for notice and takedown of dangerous products⁴⁴. Compliance with

³⁹ Masnada M., Pacciti A., Ecanova C., *op. cit.*

⁴⁰ *Ibid.*

⁴¹ *Ibid.*

⁴² *Ibid.*

⁴³ Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

⁴⁴ *Ibid.*, p. 592.

these DSA duties by a platform should make it easier to identify the manufacturer or importer. If the platform fails to disclose an identifiable producer in response to a legitimate request, one might argue that it effectively stepped out of its neutral role, potentially opening itself to PLD liability by “acting like” a distributor (though the legal threshold for that would rely on the Directive’s conditions and national implementation).

Collectively, these expansions in liable actors are designed to fulfill the Directive’s goal that there is always an EU-based entity responsible⁴⁵ for a defective product that causes harm in the EU. This protects consumers from being left remediless when dealing with global e-commerce. It also spreads the incentives for product safety across all players: manufacturers must design safe products; importers and reps must ensure the safety of what they bring in; fulfilment providers and platforms must exercise due diligence in the products they choose to handle or list (or ensure the foreign trader has an EU address for liability). Some companies may find themselves newly exposed – for example, logistics companies and online marketplaces now face potential strict liability suits. This is a notable policy choice, effectively treating certain service providers in the distribution chain as if they were producers in order to protect consumers. It will be interesting to see how these provisions are interpreted by courts – especially what it means for a platform to “create the impression of being the seller or an authorized representative”⁴⁶, which triggers liability, versus merely acting as a host. In any event, the PLD’s framework clearly aligns with the DSA’s philosophy: larger platforms, particularly, should not be able to avoid responsibility when they play a decisive role in transactions. (Indeed, many of the biggest online marketplaces are also designated “gatekeepers” under the DMA, reflecting their powerful position in the market; the PLD ensures that power comes with legal responsibility for product safety in appropriate cases.)

5. Burden of Proof and Disclosure

One of the most practical barriers victims faced under the old product liability regime was the burden of proof – specifically, proving that the product was defective and that the defect caused the damage. With increasingly complex products (like AI-driven devices, pharmaceuticals, or IoT systems), these elements can be technically challenging to demonstrate in court. The new Directive tackles this issue head-on by introducing two significant procedural innovations: (1) a mechanism for disclosure of evidence held by the defendant, and (2) a set of rebuttable presumptions that ease the claimant’s burden of proof for defect and causation under certain conditions. These changes are inspired in part by analogous developments in other areas of EU law (such as competition law damages

⁴⁵ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

⁴⁶ Masnada M., Pacciti A., Ecanova C., *op. cit.*

and consumer protection litigation) and aim to put claimants and defendants on a more equal footing⁴⁷.

A. Disclosure of Evidence. Article 8 of Directive 2024/2853 (numbered Article 9 in some commentaries referencing the proposal) establishes an obligation for defendants to disclose relevant evidence in their possession when certain conditions are met. In a product liability lawsuit, a national court can, at the request of the claimant, order the defendant to disclose evidence that is relevant to the claim **if** the claimant has already presented facts and evidence sufficient to support the plausibility of the claim⁴⁸. This is a notable shift for many EU jurisdictions that do not have broad pre-trial discovery in civil cases. The Directive essentially creates a tailored discovery mechanism: the claimant must first make a plausible case (not mere speculation – they need some indicia of defect or causation), and then the court can compel the producer (or other defendant) to provide information that could be crucial to substantiating the claim (such as internal test reports, design specifications, incident data, etc.).

Defendants are also allowed to request evidence from claimants, symmetrically, if they need it to defend themselves (for instance, if a component maker needs access to the damaged product in the claimant's possession to analyze it)⁴⁹. However, the primary rationale of this provision is addressing the *information asymmetry*: the manufacturer typically knows far more about the product's design and risks than the consumer does. By enabling courts to order disclosure, the Directive prevents manufacturers from completely hiding behind technical secrecy. There are safeguards: any disclosure order must be necessary and proportionate and must consider the legitimate interests of all parties, especially confidentiality and trade secrets⁵⁰. Indeed, the Directive explicitly references the EU Trade Secrets Directive to ensure that courts protect sensitive know-how – for example, by using confidentiality clubs or redaction as needed⁵¹.

If a defendant fails to comply with a court's evidence disclosure order, the new PLD introduces a punitive consequence: a rebuttable presumption of defectiveness can be applied in favor of the claimant in such a case⁵². In other words, if the manufacturer refuses to produce the evidence that the court has ordered, the court may presume that the product was defective (or that causation is established, depending on what the evidence pertained to), unless the defendant rebuts that presumption. This provides a strong incentive for defendants to comply with disclosure orders – non-compliance could practically hand victory to the claimant. It also prevents stonewalling; a claimant will not be prejudiced by a defendant's refusal to share information uniquely in its control. This presumption

⁴⁷ Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

⁴⁸ *Ibid.*, p. 592.

⁴⁹ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

⁵⁰ Shu Li, Beatrice Schutte, *op. cit.*, 592.

⁵¹ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

⁵² Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

for failure to disclose is codified in Article 10(2)(a) of the Directive⁵³ and is one of several presumptions discussed below.

B. Rebuttable Presumptions Easing Burden of Proof. Perhaps the most striking innovation of the Directive is found in Article 10, which sets out several rebuttable presumptions of fact that can assist claimants in proving a defect or causation. Under the traditional 1985 regime, the claimant had to prove defect, damage, and causal link with no special evidentiary presumptions (aside from whatever national evidence rules might allow, like *res ipsa loquitur analogies* in some cases). The new Directive harmonizes specific presumptions across all Member States, which is a significant step toward claimant-friendly harmonization⁵⁴. These presumptions are as follows:

- **Presumption of Defectiveness in Certain Circumstances:** Article 10(2) provides that a product shall be presumed defective (i.e., not meeting the required safety) if *any one* of the following conditions is demonstrated by the claimant: **(a)** the defendant has failed to comply with a court's disclosure order (as noted above); **(b)** the product does not comply with mandatory safety requirements that were intended to protect against the risk of the damage that occurred; or **(c)** the damage was caused by an obvious malfunction of the product under normal use⁵⁵. Condition (a) targets disclosure refusal. Condition (b) essentially means that if a product violated specific safety regulations (for example, it failed to meet an EU safety standard or was subject to a recall by a regulator), and that violation is relevant to the harm suffered, the court can presume the product was defective. This aligns with common sense: a product breaching safety laws is likely defective. Condition (c) addresses situations where a product obviously fails in a way that ordinarily it should not – for instance, a new appliance exploding or a car's brakes failing without explanation. In such cases, rather than requiring the victim to prove the precise technical defect, the law presumes defectiveness *because* such accidents don't happen absent a defect. The term "obvious malfunction" is meant to capture incidents that speak for themselves (a concept akin to *res ipsa* in tort law), and it limits it to foreseeable use or ordinary circumstances to exclude misuse scenarios.

- **Presumption of Causation:** Article 10(3) adds a presumption for the causal link between defect and damage. It states that causation shall be presumed where it is established that the product is defective and the damage is of a kind typically consistent with that defect⁵⁶. In other words, if the claimant proves (or

⁵³ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

⁵⁴ Nynke E. Vellinga, 'Rethinking Compensation in light of the Development of AI', *International Review of Law, Computers & Technology*, 38 (3) 2024, 391–412, 393.

⁵⁵ *Ibid.*, p. 392.

⁵⁶ Becker M., Bell A., Meyer H., *Product Risks Today: How the new Product Liability Directive facilitates private enforcement*, 2025, accessible on: <https://riskandcompliance.freshfields.com/post/102k71h/product-risks-today-how-the-new-product-liability-directive-facilitates-private#:~:text=with%20the%20defect%20in%20question,in%20order%20to%20allow> (last access: 02.05.2025); Shu Li, Beatrice Schutte, *op. cit.*, p. 592.

benefits from a presumption) that the product had a defect, and the harm that occurred is the kind of harm that defect would normally be expected to cause, the court can presume that the defect caused the harm. For example, if a defect in a car's airbag is established and the harm was the car occupant's injury in a crash (the kind of injury an airbag defect would contribute to), then causation between the defect and injury is presumed. This prevents defendants from exploiting uncertainties about exact causal chains when the general link is evident. It's then up to the defendant to rebut by showing the injury was actually caused by something else unrelated to the defect.

- Presumption of Defectiveness and/or Causation under High Complexity ("excessive difficulties"): Article 10(4) introduces a broad and somewhat groundbreaking presumption: if the claimant faces excessive difficulties in proving defect or causation due to technical or scientific complexity, the court may presume the product is defective or/and that it caused the damage (as relevant), *provided the claimant has demonstrated that it is likely so*⁵⁷. This essentially lowers the standard of proof to a "likelihood" when things are too complex to expect full proof. The claimant must show a plausible case that the product was likely defective or likely the cause of harm, and that obtaining more evidence is excessively difficult (perhaps because of the complexity of the AI algorithm, or multi-factor causality in a medical device's effect). If the court is convinced of those points, it can presume defect and causation. This is a powerful tool, aimed at scenarios like AI systems⁵⁸, pharmaceuticals, or other advanced tech where a victim might be at a huge disadvantage in pinpointing the exact failure. However, it is also the most debated because it verges on a partial reversal of the burden of proof. As Clyde & Co commentators noted, it's unclear when exactly a court will deem something "excessively difficult" – this will depend on national courts and could vary⁵⁹. There's concern that entire categories of products (say, all AI-driven systems or all complex medical devices) might be routinely treated under this presumption, effectively considering them presumptively defective unless proven otherwise⁶⁰. The requirement that the claimant show "likelihood" of defect or causation is also a relatively low threshold, significantly lower than "balance of probabilities" in practice⁶¹. Defendants worry this creates a near-automatic liability for cutting-edge technologies where courts might sympathize with the complexity argument⁶².

All these presumptions are rebuttable. Article 10(5) allows the defendant to rebut any of the above presumptions with evidence to the contrary⁶³. In theory,

⁵⁷ Becker M., Bell A., Meyer H., *op. cit.*, 2025.

⁵⁸ Nynke E. Vellinga, *op. cit.*, p. 393.

⁵⁹ Becker M., Bell A., Meyer H., *op. cit.*, 2025.

⁶⁰ *Ibid.*

⁶¹ *Ibid.*

⁶² Becker M., Bell A., Meyer H., *op. cit.*, 2025.

⁶³ Becker M., Bell A., Meyer H., *op. cit.*, 2025.

this preserves fairness by not making the presumptions absolute. In practice, however, once a presumption is triggered, the burden of evidence shifts to the defendant, which can be outcome-determinative if the defendant cannot muster sufficient proof. For instance, how would a producer rebut an “obvious malfunction” presumption? Perhaps by showing the product was tampered with or misused by the consumer. How to rebut the complexity presumption? Possibly by arguing the case is not as complex as claimed or by actually proving the product was not defective. The Freshfields analysis points out that these presumptions may lead to a “de facto reversal” of the burden of proof in many cases⁶⁴. From a doctrinal perspective, this is a noteworthy shift: EU product liability was traditionally strict on *liability* (no fault needed) but neutral on burden of proof, whereas now it leans towards helping the claimant prove the defect and causation.

It is worth noting that prior to this Directive, some national courts and laws in Europe had already been exploring ways to ease proof in complex product cases (for example, French courts in some drug liability cases inferred defects, and the EU Court of Justice in *Boston Scientific* (2015) allowed inference of defect for a whole product line if one product had a defect). The PLD essentially codifies a harmonized approach, ensuring all Member States will now offer at least these presumptions in product cases. This harmonization can prevent “forum shopping” and ensure a high level of consumer protection uniformly.

Trade-offs and Safeguards: While claimant-friendly, these rules try not to go too far. They do not reach the point of outright strict liability with no need to prove defect at all; the claimant still must do some work (plausibility for disclosure; triggering conditions for presumptions). They also explicitly leave it to national courts to evaluate circumstances. For instance, the “excessive difficulty” presumption is discretionary (“courts may presume”) and case-by-case. The Directive’s recitals encourage careful application to avoid automatic presumptions for broad categories without analysis⁶⁵. Over time, jurisprudence (and possibly guidance from the Court of Justice of the EU) will likely refine the boundaries of these concepts.

In sum, the PLD’s provisions on disclosure and burden of proof represent a significant development in product liability procedure. They align with a wider trend in EU law of enhancing private enforcement by empowering claimants (seen also in competition law damages directives and the Representative Actions

⁶⁴ Becker M., Bell A., Meyer H., *op. cit.*, 2025.

⁶⁵ Civatte, E., Winckler, B., O’Sullivan, J., & Dunne, S. *A new liability framework for products and AI—An update on the new EU Product Liability Directive and the proposed AI Liability Directive*, 2025, accessible on: <https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-products-and-ai/> (last access: 02.05.2025); Narayanan, S., & Potkewitz, M., *A Risk-Based Approach to Assessing Liability Risk for AI-Driven Harms Considering EU Liability Directive*, 2023, accessible on: [arXiv.https://arxiv.org/abs/2401.11697](https://arxiv.org/abs/2401.11697) (last access: 02.05.2025); Buiten M.C., ‘Product Liability for defective AI’, *European Journal of Law and Economics* 57 (2024), 239-273, p. 241.

Directive for consumers). For legal scholars, these changes raise interesting questions about the interaction with national civil procedure (which will have to accommodate these orders and presumptions) and about whether the balance struck is optimal. Producers have voiced concerns that easier litigation could lead to more claims and higher insurance costs, potentially discouraging innovation in high-tech sectors⁶⁶. On the other hand, consumer advocates argue these measures are necessary to make rights effective – a liability regime that exists on paper but is impossible to use in practice (because consumers can't prove complex defects) fails its purpose. The success of these provisions will ultimately be measured by whether they indeed improve access to justice for injured persons without unduly burdening courts or causing defensive innovation. They will certainly make product liability trials more dynamic, as parties will battle not just on substance but on whether presumptions should apply and what evidence must be disclosed.

6. Interplay with AI Liability Directive

In parallel with updating the Product Liability Directive, the European Commission in 2022 proposed a separate directive on AI-related civil liability – commonly referred to as the AI Liability Directive (AILD) – with the intent to complement the PLD in addressing harms caused by artificial intelligence systems. The rationale was that while the PLD (even as revised) covers *defective AI products* under strict liability, there could be situations where an AI system causes damage without a product defect per se, or where a fault-based claim against an AI system's provider or user is more appropriate. The proposed AILD sought to harmonize certain aspects of fault-based liability in relation to AI, ensuring that victims are not worse off in cases involving AI than in traditional cases. Understanding the interplay between the PLD and the AILD is important for a full picture of the EU's approach to AI risks – although, as will be discussed, the AILD's fate has become uncertain⁶⁷.

Scope and Purpose of the AI Liability Directive: The proposed AILD (European Commission proposal COM(2022) 496) would establish harmonized rules for non-contractual civil liability for damage caused by AI systems. Unlike the PLD, which is a strict liability regime focused on defective products, the AILD was envisioned as a fault-based regime applying to any “AI system” (as

⁶⁶ Civatte, E., Winckler, B., O'Sullivan, J., & Dunne, S., *op. cit.*, 2025; Narayanan, S., & Potkewitz, M., *op. cit.*, 2023; Hacker, P., *The European AI Liability Directives: Critique of a Half-Hearted Approach and Lessons for the Future*, 2022, Accessible on: arXiv. <https://arxiv.org/abs/2211.13960> (last access: 02.05.2025); Spindler G., *Different Approaches for Liability of Artificial Intelligence – Pros and Cons – the New Proposal of the EU Commission on Liability for Defective Products and AI Systems, – Comparative analysis of the 2022 PLD and AI Liability proposals, advocating stricter AI liability*, 2023, available on: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354468#:~:text=The%20EU%20Commission%20has%20published,for%20a%20stricter%20liability%20model, (last access: 02.05.2025); Wachter S., *op. cit.*

⁶⁷ Buiten M.C., *op. cit.*, p. 241.

defined in the AI Act) that causes damage, even if not embodied in a product or not defective. It targeted two main issues: the *opacity of AI* (“*black box*” *problem*), which makes it hard for victims to prove fault/negligence, and the *autonomy of AI*, which can make causal link determination difficult⁶⁸. To alleviate these issues, the AILD proposed two key measures: a rebuttable presumption of causality in fault-based AI claims, and a right of access to evidence about high-risk AI systems for claimants⁶⁹.

Under the proposal, if a claimant sued an operator or user of an AI system for fault (negligence), and the claimant could show that the defendant likely failed to comply with a certain duty of care (for instance, violated the AI Act’s requirements) and that this failure could reasonably be linked to the AI output that caused harm, then a presumption would arise that the defendant’s fault caused the harm⁷⁰. This is somewhat analogous to the PLD’s presumptions, but in the AI context for fault: it spares the claimant from the near-impossible task of proving the exact chain of causation inside an AI algorithm. The defendant could rebut by showing, for example, that the damage would have occurred even without their fault⁷¹. Additionally, the AILD would allow a victim to request disclosure of relevant information about high-risk AI systems (those designated as high-risk in the AI Act) from the supplier or user of that AI⁷². This complements the PLD’s disclosure by focusing specifically on AI context and including users, not just manufacturers.⁷³

Complementarity with the PLD: The PLD and AILD were designed to complement each other. The PLD covers strict liability for defective products, including AI systems considered as products (e.g., a robot or software that is defective). The AILD would cover scenarios not covered by PLD – mainly where no product is defective.⁷⁴ For instance, if an AI-powered decision system (say an AI medical diagnosis tool) is not defective per se (it meets all safety requirements and functions as intended), but a hospital or developer was negligent in its use or training (leading to harm), a victim might not succeed under PLD (no defect) but could sue under fault principles. The AILD presumptions would assist in that fault-based case. Similarly, the AILD would apply to pure software services or AI outputs not embedded in products – e.g., a financial trading AI causing pure economic loss (which wouldn’t be covered by PLD at all, as PLD doesn’t cover pure economic loss or non-material damage like discrimination harms, etc.).

⁶⁸ Nynke E. Vellinga, *op. cit.*, p. 393; Spindler G., *op. cit.*, 2023; Wachter S., *op. cit.*

⁶⁹ Civatte, E., Winckler, B., O’Sullivan, J., & Dunne, S., *op. cit.* 2025; Tiago Sergio Cabral (2020) ‘Liability and Artificial intelligence in the EU: Assessing the adequacy of the Current Product Liability Directive’, *Maastricht Journal of European and Comparative Law*, 27 (5): 615-635, p. 621-622; Hacker, P., *op. cit.*, 2022; Buiten M.C., *op. cit.*, p. 241.

⁷⁰ Narayanan, S., & Potkewitz, M., *op. cit.*, 2023.

⁷¹ Nynke E. Vellinga, *op. cit.*, p. 393; Spindler G., *op. cit.*, 2023.

⁷² Civatte, E., Winckler, B., O’Sullivan, J., & Dunne, S., *op. cit.*, 2025; Tiago Sergio Cabral, *op. cit.*, p. 621-622.

⁷³ Buiten M.C., *op. cit.*, p. 241.

⁷⁴ Nynke E. Vellinga, *op. cit.*, p. 393; Buiten M.C., *op. cit.*, p. 241.

Furthermore, the PLD and AILD both include disclosure and presumptions, but targeted differently. The PLD's disclosure is against manufacturers; the AILD's is specifically for AI and could target users. There is a conscious parallel: both aim to make litigation feasible despite complexity. In legislative debates, it was stressed that the AILD would *not* impose new bases of liability, but rather harmonize certain procedural aspects. This means a claimant still needs a cause of action under national law (like negligence) to use the AILD's tools. The PLD, by contrast, creates a direct cause of action EU-wide for defective products.

Interactions in Practice: Suppose an AI-powered autonomous vehicle causes an accident. If the accident is due to a defect in the vehicle or its AI software (e.g., a coding error, sensor failure – something making it unsafe beyond expectations), the victim can sue under PLD (strict liability against the car manufacturer). If the accident is due to no defect but perhaps the human overseer's operational error, or maybe an inherent limitation of the AI that is not a "defect" but arguably someone was negligent in deploying it in that situation, then a fault-based claim might be appropriate – for example, against the operator of the AI or the developer, depending on circumstances. The AILD would help by presuming causation if the AI likely played a role and by allowing access to logs from the AI to see if it malfunctioned. A claimant could conceivably pursue both in the alternative: a PLD claim against the manufacturer (saying the AI car was defective) and a fault claim (with AILD aids) against, say, the fleet operator for not properly monitoring the AI. The law is designed so that these avenues aren't mutually exclusive or contradictory but offer a comprehensive net of liability.

Current Status – AILD Withdrawal: As of early 2025, it's important to note that the proposed AI Liability Directive has encountered political headwinds. In the European Commission's Work Program for 2025, the Commission announced an intention to withdraw the AILD proposal due to lack of progress and concerns raised during negotiations⁷⁵. Some industry stakeholders argued that the AILD might create legal uncertainty and overlap with existing laws, and that the new PLD along with existing national tort law might suffice for AI cases⁷⁶. On 11 February 2025, the Commission listed the AILD for withdrawal, which has led to debate in Parliament and among member state⁷⁷s. Several Members of European Parliament and commentators criticized this move, suggesting that it was premature to abandon the AI-specific liability harmonization and that doing so could leave gaps or inconsistencies in protection (especially since the AI Act is moving forward, one would expect a liability counterpart)⁷⁸. As of this writing, it remains to be seen if the withdrawal will be finalized or if parts of the

⁷⁵ Duffourc M.N. (2025), *The Withdrawal of the AI Liability Directive: A Critical Reflection on AI Liability in the EU*, accessible on: <https://www.maastrichtuniversity.nl/blog/2025/02/withdrawal-ai-liability-directive-critical-reflection-ai-liability-eu> (last access: 02.05.2025).

⁷⁶ Ibid.

⁷⁷ Ibid.

⁷⁸ Ibid.

AILD will be salvaged (possibly its content could be integrated into the AI Act or other legislation).

If the AI Liability Directive does not proceed, the interplay simplifies: the PLD (as now adopted) becomes the primary EU-level tool for AI-caused harm insofar as a defect can be identified, and otherwise victims must rely on national liability laws (with no special EU presumptions for AI, unless Member States adopt their own). Some Member States might fill the gap by national legislation or by creative judicial interpretation. The PLD itself, as we have detailed, already goes a long way to address AI issues by including software/AI as products and by presumptions that help in complex cases (including those involving technical complexity – which certainly covers AI). Additionally, the AI Act (Regulation (EU) 2024/1689, expected to fully apply in 2025-2026) will impose strict obligations on providers and users of high-risk AI systems; non-compliance with those obligations could be used as evidence of fault in national tort claims. It's notable that the AILD's presumptions of causality would have kicked in if an AI provider/user violated the AI Act requirements⁷⁹. Without AILD, a national court could still potentially infer causation or fault if it sees AI Act violations, but that will depend on national doctrine.

In summary, the new Product Liability Directive and the (now uncertain) AI Liability Directive were conceived as two pillars of the EU's approach to AI accountability: one rooted in strict liability for defective products (technology-neutral but AI-inclusive), and one supplementing fault-based liability to deal with AI's peculiarities. Together, they would have ensured that whether a harm from AI was due to a product defect or some other failing, the victim had a clear path to compensation. With the PLD now in force, consumers are better protected when AI systems turn out to be defective products. If the AILD does not materialize, some risk remains that in cases of pure algorithmic decision-making harm (without a defect), victims will face the traditional difficulties in proving negligence, albeit with some help from the AI Act's transparency and record-keeping rules. European legal scholars will be watching closely how national courts handle such cases and whether the PLD's generous presumptions might sometimes be stretched to cover scenarios that AILD would have addressed (for instance, using the "excessive difficulty" presumption in a borderline case to achieve justice even absent an obvious defect). The *interplay*, therefore, might become an internal one – between the PLD and creative national fault liability – rather than between two EU directives.

7. Relationship to the DSA and DMA

The *Digital Services Act (DSA)* and *Digital Markets Act (DMA)* are two major EU regulations that came into force in 2022, fundamentally reshaping the

⁷⁹ Ibid.

legal framework for online intermediaries and large digital platforms. While their primary focus is not product liability, they establish duties and parameters that influence how online platforms operate, including in relation to the sale of products. The new Product Liability Directive intersects with these Acts in terms of platform responsibility and market fairness. This section explores how Directive 2024/2853 relates to the DSA and DMA, highlighting complementarity and potential tensions.

A. Digital Services Act (DSA) – Platform Liability and Due Diligence.

The Digital Services Act (Regulation (EU) 2022/2065) is a comprehensive framework for regulating the moderation of online content and the responsibilities of “intermediary services,” including online marketplaces. One of the DSA’s key features is that it maintains the conditional liability exemptions (safe harbors) for intermediaries that were established under the old e-Commerce Directive (2000/31/EC)⁸⁰. For online platforms like marketplaces, this typically means they are *not liable for illegal content or products sold by third parties via their platform* so long as they play a neutral, passive role (merely hosting the listings) and act expeditiously to remove or disable access when they obtain actual knowledge of illegal content⁸¹. The DSA elaborates this by saying if platforms do not “create the impression” they are the seller, and if they comply with certain transparency duties, they can avoid liability for the underlying conduct of the selling third party⁸².

Directive 2024/2853, as discussed in the Liable Actors section, introduces potential liability for online platforms *when they go beyond a neutral role* – for example, when the platform effectively controls the fulfillment or presents itself as a seller (an “active role”). There is an inherent tension to manage: the PLD cannot simply override the DSA’s safe harbor which is a directly applicable Regulation; nor does it attempt to make platforms strictly liable for every defective product sold by others on their site. Instead, the PLD and DSA regimes are designed to complement each other: the PLD attaches liability to platforms *only in situations where the platform behaves like an operator in the supply chain* (manufacturer, importer, distributor, etc.)⁸³. This is consistent with the DSA, which implies that if a platform is no longer an impartial intermediary but rather presenting itself as part of the transaction, it should not enjoy the safe harbor at all. In fact, Recital 27 of the DSA clarifies that the liability exemptions do not apply when the online platform has an active role, particularly “by presenting the content or by otherwise facilitating the transaction in a way that would lead a consumer to believe that the information or product is provided by the platform

⁸⁰ Barnes P., Kelly C., *Navigating the New EU Product Liability Directive*, 2024 accessible on: <https://www.clydeco.com/en/insights/2024/11/navigating-the-new-eu-product-liability-directive> (last access: 02.05.2025).

⁸¹ Ibid.

⁸² Ibid; V. Burgsdorff Christoph, *op. cit.*

⁸³ Barnes P., Kelly C., *op. cit.*, 2024; V. Burgsdorff Christoph, *op. cit.*, 2025.

itself.”⁸⁴ The PLD effectively leverages that principle by explicitly stating that in cases where no manufacturer/importer can be identified, a platform can be held liable if it creates the impression of being the seller or otherwise fails to identify the actual producer⁸⁵.

Moreover, the DSA imposes due diligence obligations on online marketplaces which dovetail with the PLD’s aims. Under Article 24 of the DSA, online marketplaces must collect and display certain information about the traders using their platform (a “Know Your Business Customer” requirement) and must inform consumers when they become aware that a product sold may be unsafe (this aligns with EU product safety law requirements)⁸⁶. If marketplaces comply diligently, then in theory, for every product sold, there should be an identifiable manufacturer or importer. This greatly assists the functioning of the PLD: the injured person can find the manufacturer’s identity from the platform’s records and sue the manufacturer (or importer) directly, obviating the need to sue the platform. Conversely, if a platform fails to obtain or provide that information, then the consumer is left without an obvious defendant except the platform. The PLD’s subsidiary liability for platforms in cases where no one else is identified thus provides a backstop and an incentive for platforms to follow the DSA’s mandates. We can see a policy synergy: the DSA pushes platforms to be transparent and responsible in vetting traders, and the PLD says, if you don’t, you might end up liable yourself. In practice, large platforms will likely tighten compliance with DSA obligations to avoid falling into that trap.

Another point of intersection is how the DSA’s notion of “illegal content” might encompass defective or dangerous products. The DSA mostly concerns illegal content (hate speech, IP infringement, etc.), but an unsafe product could be considered illegal to sell (for example, a banned toy due to safety reasons). The DSA’s notice-and-takedown system would require a platform to remove a listing for a product that is notified as unsafe/illegal. If a platform fails to do so and a consumer is harmed by that product, questions arise: could the platform be liable due to negligence (outside the PLD) or under the PLD as a de facto distributor? The PLD would likely regard the platform as liable if it was effectively the only entity in the EU involved, though proving the platform had the role might involve showing it didn’t act on knowledge of danger (blurring into fault). Even outside strict PLD liability, a platform that ignores its DSA obligations could face regulatory fines and perhaps national tort claims.

In summary, the relationship with the DSA is that the new PLD complements the DSA by covering the civil liability aspect that the DSA deliberately did not fully address (since DSA kept the safe harbor). The PLD’s nuanced approach to platform liability is carefully crafted not to conflict with the DSA: it does not make a compliant, purely intermediary platform liable for defects of third-party

⁸⁴ Barnes P., Kelly C., *op. cit.*, 2024; V. Burgsdorff Christoph, *op. cit.*, 2025.

⁸⁵ Barnes P., Kelly C., *op. cit.*, 2024; V. Burgsdorff Christoph, *op. cit.*, 2025.

⁸⁶ Barnes P., Kelly C., *op. cit.*, 2024; V. Burgsdorff Christoph, *op. cit.*, 2025.

products (thus preserving the safe harbor principle), but it closes the loophole of platforms acting as sellers with impunity. The DSA and PLD together ensure that online marketplaces either identify the true seller/producer or answer to the consumer themselves. This advances consumer protection in e-commerce while still protecting platforms from unreasonable burdens when they are genuinely just passive conduits.

B. Digital Markets Act (DMA) – Gatekeepers and Market Fairness.

The Digital Markets Act (Regulation (EU) 2022/1925) targets the largest digital platform companies (designated “gatekeepers”) with obligations to ensure fair competition and to curb abusive practices. At first glance, the DMA is about antitrust-like regulation and does not directly deal with product liability or consumer protection. However, the **context** it creates is relevant: many gatekeepers under the DMA (for instance, major online marketplace operators, app stores, search engines) are also players through which products reach consumers. The DMA requires gatekeepers to refrain from certain self-preferencing or tying practices and to ensure interoperability in some cases.

While the DMA doesn’t impose duties related to product safety or liability, one can argue that it complements the product liability regime by maintaining a fair environment in which competition on product safety can take place. A gatekeeper platform cannot, for example, unfairly down-rank or block third-party services that might offer better safety features (that would violate DMA’s fairness rules), which indirectly ensures consumers can access safer products or services. Additionally, the DMA’s goal of opening up platform ecosystems might mean a broader range of intermediaries and services, which in turn could affect how liability is distributed (for instance, if Apple and Google must allow alternative app stores, those alternative stores might take on roles that include liability for app defects, etc., similar to how the main platform would under PLD if acting as distributor).

Another indirect relationship is that the big gatekeeper firms are likely to be the ones most affected by the PLD’s new provisions on platforms and modifications. For example, Amazon is both a gatekeeper (under DMA) and will likely be targeted by PLD suits as a fulfilment provider or distributor if it doesn’t ensure EU-based sellers. The DMA doesn’t say anything about liability, but it ensures Amazon cannot, say, favor its own products and hide third-party risk information, etc., which overall contributes to a level playing field in assuming liability. Similarly, gatekeepers might be in the best position to absorb or manage the risks of product liability (given their resources), and the PLD effectively forces them (when no one else is in the EU chain) to do so. This aligns with the DMA’s theme that gatekeepers have a responsibility not to exploit their position in ways that ultimately harm consumers or competition. One could argue that *holding gatekeeper platforms liable for defective third-party products in certain scenarios is consistent with the idea that they cannot have all the advantage of being central marketplaces without any of the corresponding responsibility.*

It is also possible that DMA compliance could provide platforms with defenses or at least goodwill arguments in liability cases. For instance, if a gatekeeper platform is following DMA obligations and providing access/data to business users, it might claim it has done what is required on transparency and that any defect is purely on the manufacturer – though this wouldn't exempt liability under PLD if the conditions for platform liability are met, but could influence courts' view of whether the platform "acted like an operator" or just as a neutral venue.

In essence, the DMA's relationship to product liability is more contextual and indirect than the DSA's. The DMA ensures the digital market isn't monopolized or distorted, which means no single company can avoid consumer pressure to maintain safety. It also means if a gatekeeper does incur liability (under PLD or otherwise), it cannot unfairly shift that cost to others or block rivals who might be safer. There's also a philosophical alignment: both the DMA and the PLD revision are part of the EU's strategy to rein in big tech companies and ensure they operate under European standards of fairness and safety. One targets economic power, the other targets responsibility for harm. Both together send a message that large platforms operating in Europe must be accountable in various dimensions – to regulators, competitors, and consumers alike.

Finally, we note that none of the new laws (DSA, DMA, PLD) explicitly override the others; they must be read in harmony. The PLD specifically references the DSA in recitals to clarify the boundary of platform liability⁸⁷. While it does not reference the DMA (understandably, as DMA is about competition), the practical effect of DMA is simply that some platforms will have to adjust business models (for example, separating certain services) which could marginally affect how product liability risk is managed. For instance, if a gatekeeper must allow data portability and interoperability, third-party services could help consumers track product provenance or safety issues – aiding in liability claims or prevention. These are second-order effects but part of the holistic EU digital regulation ecosystem.

In conclusion, the DSA and DMA form the backdrop of rules governing platform behavior, with the DSA focusing on what platforms must do to protect users and when they are exempt from liability, and the PLD stepping in to impose liability when those conditions are not met or when the platform's role goes beyond mere conduit. The DMA ensures that these rules play out in a competitive environment without gatekeepers exploiting their position to evade new obligations. Together, these instruments demonstrate the EU's twin aims of fostering a *safe digital marketplace* and a *fair digital marketplace*. The Product Liability Directive's contribution is squarely in the safety realm, but it is carefully coordinated with the DSA's liability framework and sits conceptually comfortably

⁸⁷ Piovano, Ch., Hess Ch., *op. cit.*, p. 45; Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

alongside the DMA's ethos of accountability of powerful actors.

8. Challenges and Implementation

The adoption of the new Product Liability Directive ushers in a host of challenges and considerations for implementation. While the Directive promises a more robust and future-proof liability regime, translating its provisions into practice will require careful navigation by Member States, courts, businesses, and consumers. This section discusses some of the key challenges, including legal uncertainties in interpretation, potential impacts on innovation and insurance, and practical issues in transposition and enforcement.

A. Transposition into National Law. Member States have a two-year window (until December 2026) to transpose Directive 2024/2853 into their national legal systems⁸⁸. Despite the Directive's full harmonization intent, transposition will not be entirely uniform in practice, because national legal systems differ in procedural law and may exercise limited discretion in areas the Directive leaves open (such as determining specific procedures for evidence disclosure). One challenge is integrating the Directive's disclosure and presumption mechanisms into national civil procedure. Many EU countries historically do not have U.S.-style discovery; courts will need to be empowered (or reminded of existing powers) to order evidence disclosure consistent with Article 8 PLD. National legislators might have to clarify what "facts sufficient to support plausibility" mean in their context and set rules to prevent fishing expeditions⁸⁹. Similarly, the rebuttable presumptions of Article 10 will represent a change for judges – training and guidance may be needed so that judges know when to apply these presumptions. Some legal systems might incorporate these presumptions by an explicit statutory language, while others might rely on courts to infer them directly from the transposed text.

Another transposition issue is dealing with the interplay with existing national liability regimes. All Member States have some form of product liability law (many simply implemented the 1985 Directive into national statutes; some had parallel tort claims). Those national laws will now be replaced or amended. For example, Germany's *Produkthaftungsgesetz* and France's Civil Code provisions on product liability will be updated to reflect the new definitions (like including software as product), new liable parties, etc. This is mostly straightforward, but some countries had taken divergent approaches on points the 1985 Directive left optional. For instance, Spain and Finland had opted *not* to allow the development risks defense, whereas the UK (when it was under the EU) and Germany allowed it in full or part⁹⁰. Now, the Directive mandates a harmonized ap-

⁸⁸ See also V. Burgsdorff Christoph, *op. cit.*

⁸⁹ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

⁹⁰ Piovano, Ch., Hess Ch., *op. cit.*, p. 65; Wybitul T., Sikora T., *New EU Product Liability Directive*

proach to development risks (i.e., generally allowing the defense except for specified circumstances). Those Member States will have to introduce the defense into their law, which could be politically sensitive – consumer advocates in those countries may resist, seeing it as a step backward in consumer protection. However, because it's an EU full harmonization, they have no flexibility on that point. We might see some Member States emphasizing the exceptions (software, etc.) to narrow the defense as much as allowed, or providing for strict judicial scrutiny when a producer invokes it.

Conversely, Member States will remove things like the €500 property damage threshold and the exclusion of business property damage, which were in the old directive. Now *any* level of property damage is compensable and not limited to private use⁹¹. This is a clear improvement for claimants (no small claims bar), but for insurers and businesses it means more potential low-value claims to handle. Implementing this could lead to a higher volume of claims for minor damage (e.g., a defective kitchen appliance slightly damages a countertop – previously under €500 not claimable via PLD, now it is). National rules might need to adapt their small claims procedures or encourage alternative dispute resolution for very low-value claims to avoid court overload.

One practical implementation challenge is creating the infrastructure for evidence disclosure. Courts must be ready to handle sensitive information, possibly using confidentiality measures. This may be new for some civil law courts. Additionally, businesses might need to adjust record-keeping: knowing that they could be ordered to disclose internal data on products years later (remember the long liability period can be up to 10 or even 25 years for latent injuries⁹²), manufacturers should maintain archives of design documents, test results, and incident reports in a retrievable form. Member State lawmakers might also decide which courts (e.g., specialized chambers or commercial courts) should handle such cases given the technical complexity – some might channel product liability claims to courts that already handle complex litigation.

B. Interpretative Uncertainties. Despite the detailed provisions, some terms in the Directive will likely require judicial interpretation, potentially by the Court of Justice of the EU (CJEU) to achieve uniformity. We have already flagged “excessive difficulties... due to technical or scientific complexity” as one such term regarding the burden of proof presumption⁹³. National courts will have to determine on a case-by-case basis when to apply this. One can foresee early cases where defendants argue that claimants are too quick to invoke complexity

Comes into Force, Latham & Watkins Privacy & Cyber Practice 23 December 2024 | Number 3319, accessible on: <https://www.lw.com/en/offices/admin/upload/SiteAttachments/New-EU-Product-Liability-Directive-Comes-Into-Force.pdf#:~:text=of%20proof,Expanded%20Product%20Definit> on (last access: 02.05.2025).

⁹¹ Piovano, Ch., Hess Ch., *op. cit.*, p. 50, 75; Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125.

⁹² Wybitul T., Sikora T., *op. cit.*, 2024.

⁹³ Masnada M., Pacciti A., Ecanova C., *op. cit.*; Piovano, Ch., Hess Ch., *op. cit.*, p. 77.

without trying to prove their case, and claimants arguing that any AI or pharmaceutical case should qualify. Over time, precedent will likely set some thresholds (perhaps complexity in the sense of requiring expertise beyond normal human ken, etc.). The CJEU might be asked via preliminary reference: for example, “How should ‘excessive difficulty’ in Article 10(4) be assessed? Must the claimant first attempt standard proof, or can the presumption be applied from the outset?”

Another likely point of contention is the liability of online platforms. The Directive text states platforms can be liable “under certain conditions” when acting beyond mere intermediaries⁹⁴. What exactly creates the “impression” of being a seller (as referenced in Hogan Lovells commentary) and thus voids the safe harbor? Is offering a payments service or guarantee enough to consider the platform as a direct party? Some platforms provide warranties or guarantees that make them appear responsible. National courts, guided by DSA language and the PLD, will draw these lines. There may also be arguments about conflict of laws: e.g., if a platform is protected by DSA at EU level, can national implementation of PLD impose strict liability? The likely resolution is as discussed: no conflict because PLD only targets when DSA doesn’t apply. But we could see initial legal friction if, say, a platform is sued and claims immunity under DSA Article 6 (safe harbor), and the claimant says, “But under PLD you’re a distributor.” Courts might need to reconcile those; perhaps the CJEU will clarify that the PLD as a later measure effectively defines when the safe harbor doesn’t apply in product cases.

Additionally, the concept of “significant modification” of a product (making one a new manufacturer) could be tested. What changes qualify as substantial enough? Software updates could be contentious – e.g., if a third party hacks a product to change its functionality, are they a modifier (likely yes, but then they probably won’t be identifiable or solvent)? Or if a user just swaps out a part, they wouldn’t normally be “manufacturing,” but a business refurbishing multiple units might be. The Directive recitals probably give examples, but national courts will have to delineate so that, for instance, repair shops know if they risk liability.

9. Impact on Innovation and Business Practices

A. The business’ feedback. The new Directive has elicited mixed reactions regarding its impact on innovation and business. On one hand, it increases exposure to liability for producers and even peripheral actors (like fulfilment providers). This likely means higher insurance costs for product liability coverage. Insurers will assess that claims might be easier to bring (due to presumptions) and

⁹⁴ Wybitul T., Sikora T., *op. cit.*, 2024; Piovano, Ch., Hess Ch., *op. cit.*, p. 85.

that more entities are insured (logistics companies might now buy product liability insurance). Particularly, the extension of liability up to 25 years for certain latent personal injuries (for example, if someone is harmed by a product but the harm is discovered much later, as with some medical implants) means underwriters have to consider a longer “tail” risk. Businesses dealing in AI and emerging tech may see insurers raising premiums or asking for specific risk mitigation (such as rigorous record-keeping and compliance with standards to have defenses).

Some industry voices worry that the combination of strict liability and easier proof will create a litigation-friendly environment that could “discourage innovation,” especially in AI and pharmaceuticals where unknown risks are a fact of development⁹⁵. The development risk defense being mandatory might alleviate some concerns (since producers know they have that escape for unknown risks), but the carve-outs (software, etc.) mean for many tech products that defense isn’t available. Firms might respond by investing more in testing and compliance (which is a positive outcome for safety) or by hesitating to introduce products in the EU until they are very certain of safety (which could slow deployment of new tech in EU relative to elsewhere). The EU consciously accepts a bit of that trade-off, prioritizing safety and consumer trust – which, arguably, in the long run also benefits innovation by avoiding scandals that erode public confidence.

Small and medium enterprises (SMEs) might face challenges, as they have fewer resources to handle litigation or compliance. The Directive, however, does not differentiate by company size (except that micro-enterprises might rarely be fulfilment providers for global trade). One might see increased reliance by SMEs on insurance and perhaps on contractual agreements – e.g., a small manufacturer might contractually require its authorized representative to share liability or a supplier to indemnify them if a component is defective. The Directive prohibits contractual exclusion of liability towards the injured person⁹⁶, but it doesn’t prevent internal indemnity deals among companies⁹⁷. Indeed, it states liability cannot be excluded by contract vis-à-vis the victim, ensuring, for example, a platform cannot make a consumer waive their rights. But business-to-business contracts allocating the final financial burden are still possible. We may see those adjustments as companies in a supply chain negotiate who will bear the risk and cost if something goes wrong.

B. Enforcement and Litigation Landscape. From the perspective of injured persons and their advocates (lawyers, consumer organizations), the new PLD offers more levers to succeed in claims. However, awareness and effective use of these tools will be essential. Judicial training and information for legal professionals will be needed so that, for example, a judge in a Member State who’s

⁹⁵ V. Burgsdorff Christoph, *op. cit.*

⁹⁶ *Ibid.*

⁹⁷ Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125; Merryl Hervieu, *op. cit.*, Sylvie Gallage-Alwis and Gaetan de Robillard, *op. cit.*

never had discovery can manage a disclosure order, or a claimant's lawyer knows to plead the conditions for presumptions (rather than just asserting defect). The European Commission and national authorities will likely engage in outreach.

Moreover, the Representative Actions Directive (EU) 2020/1828 is coming into force around the same time (by end of 2022, Member States needed to transpose it). This allows qualified entities (like consumer associations) to bring collective actions for consumers, including for damages in mass harm situations. The PLD could see synergy with that: if a defective product injures many consumers, a representative action could be brought on their behalf. The new PLD's rules would apply in such a lawsuit. For instance, if a certain model of appliance has a defect causing fires, a consumer organization could sue the manufacturer on behalf of a class of consumers for damages, and they could invoke the presumptions (like if it obviously malfunctioned causing fire, defect is presumed). This is a new avenue that was not effectively available under the old regime (some countries had class actions, many didn't). So, one challenge is how courts will handle collective redress in product liability – something relatively novel in Europe. It could lead to more large-scale settlements or court rulings holding companies liable to many claimants at once.

Regulatory coordination is another factor: product safety regulators (under the General Product Safety Regulation 2023/988 and various sectoral laws) will continue to do market surveillance and order recalls of dangerous products. Information from regulators (like recall notices, safety test failures) could serve as evidence in product liability litigation. Conversely, the outcomes of product liability cases might flag issues to regulators. Ideally, there should be feedback loops – something the Directive doesn't explicitly provide, but national practice could develop. If a court finds a product defective and causing harm, that info could be passed to market surveillance authorities to take broader action, protecting others. Also, the Directive's requirement for the Commission to set up a database of judgments (Article 17 of the Directive mentions a public database of relevant judgments)⁹⁸, will over time create a resource where trends and precedents can be tracked across the EU. This can help identify problem products and also help harmonize interpretation as courts may look to see how others decided similar cases.

C. Balancing Consumer and Industry Interests. Critics from industry point out the potential for what they term “over-deterrence” – if companies are too fearful of liability, they might hold back beneficial products, or pass on increased costs to consumers. Consumer advocates, on the other hand, argue that the prior regime under-compensated victims and allowed companies to externalize costs of defects. The new Directive tries to balance these by measures like the development risk defense (to not punish unknowable risks) and by keeping the

⁹⁸ Becker M., Bell A., Meyer H., *op. cit.*, 2025; Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*, p. 109–125; Merryl Hervieu, *op. cit.*; Sylvie Gallage-Alwis and Gaetan de Robillard, *op. cit.*

scope to material damages (including data) and personal injuries – it does not cover pure economic loss or privacy violations, etc., which some had discussed but ultimately remained outside, as those are handled by other laws (e.g., GDPR for data breaches). Non-material harm (like pain and suffering) is only compensable if national law allows for it for personal injury (most do), which is consistent with existing practice⁹⁹. So psychological harm from injury is in, but say, loss of enjoyment or fear without physical injury remains generally out.

One challenge for consumers will be proving damages, especially data loss. The Directive says data destruction should be compensated including the cost of recovery¹⁰⁰. How do we quantify, for example, the loss of a family photo archive versus the cost to maybe attempt recovery? These are new frontiers for courts, and claimants will need expert evidence to value data. Some jurisdictions may have difficulty with the idea of compensating data loss if there's no direct economic value; others may analogize it to property.

10. Conclusions

The EU will undoubtedly monitor the effects of the new Directive. The text likely includes a review clause after some years. If it turns out that, for example, certain presumptions are not working as intended (perhaps courts rarely use the complexity presumption, or conversely, it's used too freely), the Commission could issue guidance or propose tweaks in the future. The fate of the AI Liability Directive also hangs in the balance: if not enacted now, perhaps elements of it will resurface or be integrated into national laws. Additionally, as technology evolves (e.g., biotech, IoT, etc.), the product liability regime might need further calibration. For instance, if “services” (like pure services causing harm) become a bigger issue, there might be pressure in future to extend strict liability to certain services, which the PLD still doesn't do except where a service is part of a product's functioning.

In implementing the PLD, stakeholders such as business associations and consumer groups will likely publish guidelines or best practices. Manufacturers may develop internal protocols for compliance: for example, ensuring any software updates have rigorous safety checks (knowing that lack of an update can't be used as a defense, they must supply updates responsibly or face liability). Platforms will refine their terms with sellers to ensure they get the info needed to identify manufacturers (perhaps even requiring foreign sellers to have EU importers or else they won't list them, to avoid being saddled with liability).

Finally, a challenge worth noting is the transitional period: products placed on the market before 9 December 2026 remain governed by the old rules, even if litigation happens after that date¹⁰¹.

⁹⁹ Becker M., Bell A., Meyer H., *op. cit.*, 2025.

¹⁰⁰ Nynke E. Vellinga, *op. cit.*, p. 393.

¹⁰¹ Wybitul T., Sikora T., *op. cit.*, 2024; Jacquemin, Z., *op. cit.*, p. 126–139; Koch, B. A., *op. cit.*,

Bibliography

1. Barnes, P. & Kelly C., *Navigating the New EU Product Liability Directive*, 2024 accessible on: <https://www.clydeco.com/en/insights/2024/11/navigating-the-new-eu-product-liability-directive> (last access: 02.05.2025).
2. Becker M., Bell A., Meyer H., *Product Risks Today: How the new Product Liability Directive facilitates private enforcement*, 2025, accessible on: <https://riskandcompliance.freshfields.com/post/102k71h/product-risks-today-how-the-new-product-liability-directive-facilitates-private#:~:text=with%20the%20defect%20in%20question,in%20order%20to%20allow> (last access: 02.05. 2025).
3. Buiten, M.C., 'Product Liability for defective AI', *European Journal of Law and Economics* 57 (2024), 239-273.
4. Burgsdorff Christoph, V., *Increased liability due to the new EU Product Liability Directive: what does this mean for the medical and pharmaceutical industry?*, accessible on: <https://www.ibanet.org/increased-liability-eu-product-directive> (last access: 02.-5.2025).
5. Cabral, Tiago Sergio (2020) 'Liability and Artificial intelligence in the EU: Assessing the adequacy of the Current Product Liability Directive', *Maastricht Journal of European and Comparative Law*, 27 (5): 615-635.
6. Civatte, E., Winckler, B., O'Sullivan, J., & Dunne, S. *A new liability framework for products and AI-An update on the new EU Product Liability Directive and the proposed AI Liability Directive*, 2025, accessible on: <https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-product-s-and-ai/> (last access: 02.05.2025).
7. Duffourc, M.N., *The Withdrawal of the AI Liability Directive: A Critical Reflection on AI Liability in the EU*, 2025, accessible on: <https://www.maastri chtuniversity.nl/blog/2025/02/withdrawal-ai-liability-directive-critical-reflecti on-ai-liability-eu> (last access: 02.05.2025).
8. Faure, Michael G., 'Product Liability and Product Safety in Europe: Harmonization or Differentiation?' *Kyklos*, 53 (2000) 4: 467-508.
9. Förster, Sven & Dardan Gashi (2024), *The EU's new Product Liability Directive (from a German perspective)*, <https://www.clydeco.com/en/insights/2024/04/the-eu-s-new-product-liability-directive>.
10. Gallage-Alwis, Sylvie and Gaetan de Robillard, *Product regulation and liability in France* (Signature Litigation, 2024) accessible on: <https://www.signatureliti gation.com/sylvie-gallage-alwis-and-gaetan-de-robillard-discuss-product-regu lation-and-liability-in-france-in-lexology> (last access: 02.05. 2025).
11. Hacker, P., *The European AI Liability Directives: Critique of a Half-Hearted Approach and Lessons for the Future*, 2022, Accessible on: <https://arxiv.org/abs/2211.13960> (last access: 02.05.2025).
12. Hervieu, Merryl, *Point sur la nouvelle directive européenne (UE) 2024/2853 relative à la responsabilité du fait des produits défectueux*, Dalloz, 2025, acces- sible on: <https://actu.dalloz-etudiant.fr/a-la-une/article/point-sur-la-nouve lle-di rective-europeenne-ue-20242853-relative-a-la-responsabilite-du-fait-de/>

- h/256e035c15335593d9c1bb38f7809c83.html?utm_source=chatgpt.com (last access: 02.05.2025);
13. Heydari, Tahoori, 'A Review of the Product Liability Directive and the Proposal for a Directive of Liability for Defective Products', (2024) *Al-Zaytoonah University of Jordan Journal for Legal Studies*, Special Issue, 962-971, 968-969.
14. Jacquemin, Z., 'Product Liability Directive: Disclosure of Evidence, the Burden of Proof and Presumptions', *Journal of European Tort Law*, 2024, 126–139.
15. Koch, B. A., 'Product Liability on the Way to the Digital Age', *Journal of European Tort Law*, 2024, 109–125.
16. Li, Shu and Beatrice Schutte, 'The Proposal for a Revised Product Liability Directive: The Emperor's New Clothes? (2023) *Maastricht Journal of European and Comparative Law* 30 (5): 573-596.
17. Mann, Lawrence C. and Peter R. Rodrigues, 'The European Directive on Products Liability: The Promise of Progress? (1988) 18 391 *Georgia Journal of International and Comparative Law*, 391-426.
18. Masnada M., Pacciti A. & Ecanova C., *EU introduces comprehensive digital-era Product Liability Directive*, 2024, accessible on: <https://www.hoganlovells.com/en/publications/eu-introduces-comprehensive-digitalera-product-liability-directive#:~:text=when%20online%20platforms%20function%20solely,once%20again%2C%20consistently%20with%20the> (last access: 02.05.2025).
19. Narayanan, S., & Potkewitz, M., *A Risk-Based Approach to Assessing Liability Risk for AI-Driven Harms Considering EU Liability Directive*, 2023, accessible on: arXiv. <https://arxiv.org/abs/2401.11697> (last access: 02.05.2025).
20. Piovano, Ch. & Hess Ch., *Das neue europäische Produkthaftungsrecht – EU-Produkthaftungsrichtlinie (ProdHaftRL)*. Nomos Verlagsgesellschaft, 2024, p. 25.
21. Rohrießen, B., *Die EU-Produkthaftungs-RL 2024: Der „final compromise text“ Verschärfte Produkthaftung plus Product Compliance-Pflichten im Zeichen von Digitalisierung, KI und Globalisierung*, accessible on: https://www.nomos.de/wp-content/uploads/2024/02/NL-Product-Compliance_Februar-24_Zeitschriften-Archiv_Rohrssen_GesamtPDF.pdf#:~:text=mehrfach%20verschärft,dem%20Parlament%20übersandt%2C%20nachdem (last access (02.05.2025).
22. Spindler, G., *Different Approaches for Liability of Artificial Intelligence – Pros and Cons – the New Proposal of the EU Commission on Liability for Defective Products and AI Systems, – Comparative analysis of the 2022 PLD and AI Liability proposals, advocating stricter AI liability*, 2023, available on: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354468.#:~:text=The%20EU%20Commission%20has%20published,for%20a%20stricter%20liability%20model, (last access: 02.05.2025).
23. Triaille, J., 'The EEC Directive of July 25, 1985, on Liability for Defective Products and Its Application to Computer Programs' (1993) *Computer Law and Security Report*, 215.
24. Veldt, Gitta, 'The New Product Liability Proposal – Fit for the Digital Age or in Need of Shaping up?', *EuCML* 1 2023, 24-31.
25. Vellinga, Nynke E., 'Rethinking Compensation in light of the Development of AI', *International Review of Law, Computers & Technology*, 38 (3) 2024, 391-412.

26. Wachter, S. „Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the EU, the US, and Beyond”, *Yale Journal of Law & Technology*, vol. 26, no. 3, 2024, accessible on: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4924553 (last access: 02.05. 2025).
27. Wybitul, T. & Sikora T., *New EU Product Liability Directive Comes into Force*, Latham & Watkins Privacy & Cyber Practice 23 December 2024 | Number 3319, accessible on: <https://www.lw.com/en/offices/admin/upload/SiteAttachments/New-EU-Product-Liability-Directive-Comes-Into-Force.pdf#:~:text=of%20proof,Expanded%20Product%20Definition> on (last access: 02.05.2025).

Digital Rights in the Age of Artificial Intelligence: Challenges and Perspectives

Lecturer Aurel Octavian PASAT¹

Abstract

As AI technologies become deeply embedded in society, new legal challenges related to the collection and use of personal data arise, as well as risks associated with mass surveillance and digital censorship. The article explores the growing impact of artificial intelligence (AI) on digital rights, emphasising issues such as privacy, data access and freedom of expression. It also emphasises the need to strike a balance between technological innovation and the protection of citizens' fundamental rights, through a comparative analysis of different legal systems. It takes a look at the data protection legislative framework in the European Union (GDPR) versus that in the United States, examining emerging challenges and opportunities. Relevant case studies are used to illustrate where regulations can be implemented effectively or where they are insufficient, suggesting possible solutions and future directions.

Keywords: digital rights, artificial intelligence, fundamental rights, legal systems.

JEL Classification: K24, K38

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/08>

Please cite this article as:

Pasat, Aurel Octavian, „Digital Rights in the Age of Artificial Intelligence: Challenges and Perspectives”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 144-161.

1. Introduction

Digital rights are a set of fundamental rights and freedoms that are applicable in the digital environment. They are derivatives of human rights, adapted to protect individuals in the face of the technological challenges and risks of the digital age. In the context of Artificial Intelligence, digital rights take on new dimensions, requiring a complex and dynamic approach to keep pace with rapid developments in technology.

The main digital rights in the era of Artificial Intelligence are:

¹ Aurel Octavian Pasat - Cross-border Faculty, „Dunarea de Jos” University of Galati, Romania, ORCID: 0000-0002-7239-0808, aurel.pasat@uagl.ro.

- *Right to privacy and personal data protection.* This right protects individuals' personal information from unauthorised collection, processing and sharing. In the age of AI, the emphasis is on user control over their data and transparency of automated data processing. AI algorithms can analyse and predict users' behaviour based on personal data, raising issues of surveillance and privacy violations. Facial recognition technologies and targeted advertising use personal information, often without clear consent or full understanding from the user.

- *Right to information and algorithmic transparency.* Users have the right to understand how their data is collected, used and processed by AI systems. Algorithmic transparency refers to the ability of individuals to know the principles and methods by which AI arrives at certain decisions or recommendations. Opaque algorithms or so-called 'black boxes' are difficult to understand even for experts, which complicates monitoring and ensuring fairness. Social media platforms and search engines use complex algorithms to monitor content, but users are often unaware of the criteria or data used for these decisions.

- *Right to freedom of expression and avoiding algorithmic censorship.* Freedom of expression is also protected in the digital environment, but AI used to moderate content can create risks of unwarranted censorship or manipulation of information. Algorithms that moderate content can have biases that lead to incorrect removal of content or the promotion of biased opinions. This can affect a plurality of opinions and access to information. Automated decisions by moderation algorithms on social platforms can lead to the removal of content that is deemed 'dangerous' or 'offensive' but which does not clearly violate community rules.

- *Right to justice and protection against automated decisions.* This right ensures that individuals are protected from automated decisions that affect their lives, without human intervention or the ability to challenge those decisions. It also covers the right to be informed and to understand the impact of AI algorithms on personal rights. Automated AI decisions are already being used in critical sectors such as finance, the judiciary or in the employment process, and the lack of human intervention or the possibility of challenging these decisions can have serious consequences. Credit scores automatically calculated by AI can influence a person's ability to obtain a loan, and users may not have effective means to challenge or understand the decision.

- *Right to digital security and protection against abuse.* This right protects users from cyber attacks, data breaches and misuse of digital information. AI can be used to identify and prevent such threats, but it can also facilitate advanced attacks. As AI technologies become more advanced, cybersecurity risks become more complex, requiring advanced protection and response measures. AI can be used to launch sophisticated phishing attacks or analyse vulnerabilities in security networks.

2. The Importance of Protecting Digital Rights in the AI Era

Digital rights are becoming essential in an increasingly interconnected and automated world. Everyone leaves a digital data trail on a daily basis; perhaps by buying a coffee using a reward account or by using an electronic toll collection system². AI technologies, while beneficial in many ways, have the potential to profoundly affect users' digital lives and privacy. Thus, it is important that laws constantly evolve to protect these fundamental rights and to ensure a balance between innovation and respect for democratic principles. This emphasises the importance of proactive regulation and global collaboration between lawyers, technology experts, ethicists and policy-makers. Digital rights are thus at the centre of contemporary debates on technology and society, even if Peacock³ points out, there is still debate over whether 'access to the Internet is a human right in and of itself, part of already-existing freedom of expression guarantees, or not a right at all'.

2.1. Privacy and Data Protection: the Impact of AI on Privacy

In the digital age, AI plays a significant role in transforming the way personal data is collected, processed and used. Advanced AI algorithms are capable of analysing massive volumes of data with unparalleled speed and accuracy, and this has major implications for users' privacy. While AI brings benefits, such as personalising digital experiences and improving service efficiency, these technological capabilities also create serious privacy risks. However, the privacy paradox states that the information disclosure of Internet users is problematic; although many people are concerned about their privacy online, they still share plenty of personal information on the web.⁴

Personal data is collected and processed by AI as follows:

- *Personalisation of content*: AI algorithms are widely used by digital platforms to personalise the content delivered to users, such as recommendations of movies, articles or advertisements. To do this, AI collects and analyses detailed

² Chih-Liang Yeh (2018), „Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers”, *Telecommunications Policy*, Vol. 42, Issue 4, pp. 282–292, <https://doi.org/10.1016/j.telpol.2017.12.001>.

³ Anne Peacock (2019). *Human rights and the digital divide*. London, Routledge, p. 4, <https://doi.org/10.4324/9781351046794>.

⁴ Tobias Dienlin, Philipp K. Masur, Sabine Trepte (2023). „A longitudinal analysis of the privacy paradox”. *New Media & Society*, 25(5), 1043-1064. <https://doi.org/10.1177/14614448211016316>; Alessandro Acquisti, Jens Grossklags (2003), *Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior*, UC Berkeley 2nd Annual Workshop on “Economics and Information Security”, available at: https://infoecon.net/workshop/downloads/2003/pdf/Final_session6_acquisti.pdf, accessed on 07.05.2025.

data about user preferences and behaviours, including browsing history, geographic location, social interactions and more. This intensive data collection raises concerns about the constant surveillance of users and the lack of transparency about how their data is used.

- *Targeted advertising*: The advertising industry uses AI to analyse and predict users' consumer preferences. Algorithms can identify behavioural patterns and create detailed profiles for each user to deliver the most relevant ads. This targeting capability may seem harmless at first glance, but it can reveal intimate aspects of personal lives, for example health, political or religious beliefs, and can be considered invasive.

- *Digital surveillance*: AI technologies are used by governments and private organisations to monitor people's online and offline activities. For example, facial recognition and real-time video stream analysis are used for mass surveillance. These technologies can violate the right to privacy, especially when there is no solid legal framework regulating the limits of their use. In some cases, surveillance technologies are used to monitor citizens without their consent, raising serious human rights concerns.

Major privacy challenges in the AI era are:

- *Lack of transparency*: One of the biggest obstacles to protecting privacy is the opacity of AI algorithms. Most users are unaware of how their data is collected and used and who has access to it. Moreover, the complex nature of AI algorithms makes it difficult to understand decision-making processes, complicating the ability to trace accountability and ensure fair data processing.

- *Profiling and discrimination*: Using personal data to create predictive profiles can lead to discrimination, exclusion and may affect freedom of thought⁵. For example, AI may exclude individuals from employment or credit opportunities on the basis of criteria that are not obvious and may be profoundly unfair. The data collected and processed may reveal sensitive information, exposing users to risks of discrimination based on gender, race, religion or sexual orientation.

- *Government surveillance*: Some countries use AI to monitor citizens for national security or law enforcement purposes. This surveillance can be highly intrusive and can have a negative impact on individual liberty, especially in the absence of strict regulations to limit abuses. Privacy thus becomes increasingly vulnerable in a context of pervasive digital surveillance.

Case study: The impact of facial recognition on privacy. A notable example of how AI can affect privacy is the widespread use of facial recognition technology. In cities such as London and San Francisco, AI-enabled surveillance cameras are being used to monitor public spaces. While this technology is being touted as a means to improve public safety, it raises serious questions about the sanctity of privacy. For example, there are cases where citizens' biometric data

⁵ Simon McCarthy-Jones (2019) „The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century”, *Frontiers in Artificial Intelligence*, Vol. 2, <https://doi.org/10.3389/frai.2019.00019>.

has been stored without consent or used to track political activists, undermining freedom of association and expression.

3. Legal Framework in the EU: GDPR

The General Data Protection Regulation (GDPR) is one of the world's most comprehensive and stringent regulations on the protection of personal data. Adopted by the European Union in 2018, the GDPR provides a robust legal framework aimed at protecting the privacy and personal data of European citizens. The regulation sets high standards for how organisations collect, manage and use personal data, imposing clear user rights and severe penalties for violating those rights.

3.1. Fundamental Rights Guaranteed by GDPR

GDPR provides a number of fundamental rights to EU citizens, designed to protect their privacy and give them control over their personal data. These rights include:

- *Right to information*: Organisations are obliged to provide users with clear and understandable information about how their data is collected and used. Users must be informed about the purposes of processing, the duration of data storage and the rights they have under the GDPR.

- *Right of access*: Citizens have the right to access the personal data an organisation holds about them. This includes obtaining information about the categories of data processed, the purpose of the processing and any third parties to whom the data has been disclosed.

- *Right to rectification*: Users may request the correction of inaccurate or incomplete personal data. Organisations must comply with such requests without undue delay.

- *Right to erasure* ('Right to be forgotten'): Citizens can request deletion of personal data in certain circumstances, such as when the data is no longer necessary for the purposes for which it was collected or when users withdraw their consent.

- *Right to restriction of processing*: Users can ask to restrict the processing of their personal data in certain circumstances, for example, if the accuracy of the data is contested.

- *Right to data portability*: the GDPR allows users to receive the personal data they have provided to an organisation in a structured, commonly used and machine-readable format. They can also request the transfer of this data to another organisation.

- *Right to object*: Citizens have the right to object to the processing of their personal data for direct marketing, scientific research or statistical purposes, depending on the circumstances.

- *Rights related to automated decisions and profiling*: the GDPR guarantees protection against decisions based solely on automated processing (including profiling) that have a significant impact on users. Citizens have the right not to be subject to such automated decisions if they have not consented to them or if they are not necessary for the performance of a contract.

The GDPR includes severe penalties for companies that fail to comply with its requirements, and EU regulators have imposed significant fines to ensure compliance. These fines can reach up to €20 million or 4% of a company's annual global turnover, whichever is higher. Here are some important examples:

- *Google*: In 2019, Google was fined €50 million by the French data protection authority, CNIL (Commission Nationale de l'Informatique et des Libertés), for breaching GDPR. The CNIL found that Google failed to provide users with clear and transparent information about its data processing policy and failed to obtain valid consent for the personalisation of ads. This sanction emphasised the importance of transparency and consent in the management of personal data.

- *Facebook (Meta Platforms Inc.)*: In 2021, Facebook was fined €265 million by Ireland's Data Protection Commission (DPC) for a massive data breach that exposed the personal information of more than 533 million users worldwide. The leaked data included phone numbers, email addresses and other sensitive information, and the investigation found that Facebook failed to comply with GDPR data security requirements.

- *British Airways*: In 2020, British Airways was fined £20 million for a data breach that affected the personal information of around 400,000 customers. The investigation found that the company failed to take adequate measures to protect personal data in breach of GDPR regulations.

H&M: In 2020, H&M was fined €35m for unlawfully collecting and storing excessively detailed information about its employees. The investigation found that H&M managers in Germany recorded details about employees' personal lives, such as health issues and religious beliefs, violating their privacy and fundamental rights.

The implementation of the GDPR has forced companies to become more accountable and transparent in their handling of personal data, leading to better protection of the fundamental rights of European citizens. However, significant challenges remain, such as ensuring compliance among global companies and striking a balance between privacy protection and technological innovation.

The GDPR is a gold standard example of data protection, inspiring other countries to adopt similar laws, and is a benchmark for regulating AI and its impact on privacy.

4. US and Data Protection

In the United States, the protection of personal data is regulated in a fragmented manner, with no homogenous federal legislation similar to the General

Data Protection Regulation (GDPR) in the European Union. The US approach to data privacy and security is mainly influenced by state-specific regulations and a number of sector-specific laws at the federal level. This complex structure creates a number of challenges and criticisms, especially compared to the strict standards imposed by GDPR in Europe.

Lack of Homogeneous Federal Legislation. One of the key features of data protection regulation in the US is the lack of a single, comprehensive and uniform federal legislative framework covering all aspects of personal data privacy. Instead, data protection is ensured through a combination of federal laws, state regulations and industry-specific rules. This decentralised approach can lead to inconsistencies and gaps in protecting citizens' privacy.

State laws. A notable example of state-level regulation is the California Consumer Privacy Act (CCPA), which is one of the strictest and most comprehensive data privacy laws in the US. Passed in 2018 and implemented in 2020, it has been called the US equivalent of GDPR⁶. The US corporations then promptly put resources to lobby against the California law, as the industry was concerning that the CCPA would become a de facto national standard.⁷

The CCPA gives California citizens extensive rights over their personal data, including the right to know what information is being collected, to request deletion of data, and to refuse the sale of personal data. The CCPA is seen as a benchmark for other states looking to introduce similar regulations.

Federal sectoral laws. Instead of a single federal data protection law, the US relies on laws regulating privacy in specific sectors. For example:

- *Health Insurance Portability and Accountability Act (HIPAA):* protects the privacy of patient health data.

- *Children's Online Privacy Protection Act (COPPA):* Protects children under 13 online.

- *The Gramm-Leach-Bliley Act (GLBA):* Regulates the privacy of consumer financial information.

This sector-specific approach may leave certain categories of personal data unprotected or weakly protected, depending on the nature of the activities or the jurisdiction of the application.

Criticisms of Weaker Data Protection Compared to the EU. The US regulatory model is often criticised for offering weaker protection of personal data compared to the rigorous standards imposed by the GDPR in the EU. There are several issues underlying these criticisms:

⁶ Jeeyun (Sophia) Baik (2020), „Data privacy against innovation or against discrimination?: The Case of the California Consumer Privacy Act (CCPA)”, *Telematics and Informatics*, Vol. 52, 101431, <https://doi.org/10.1016/j.tele.2020.101431>.

⁷ Matti Minkkinen, (2019). „Making the future by using the future: A study on influencing privacy protection rules through anticipatory storylines”. *New Media & Society*, 21(4), 984-1005. <https://doi.org/10.1177/1461444818817519>.

- *Lack of general protection.* Unlike the GDPR, which provides protection for all EU citizens' personal data, regardless of sector, US law is fragmented and only applies in certain contexts. For example, many of the rights offered by the GDPR, such as the right to erasure ('Right to be forgotten') or the right to data portability, are not universally recognised in the US.

- *Consent and transparency:* the GDPR imposes strict requirements to obtain informed consent from users before collecting and processing personal data. In the US, many companies may collect data without explicit consent, using general privacy clauses that are often difficult to understand or ambiguous. This has led to abuses and use of data without the full knowledge of users.

- *Selling and monetising data:* Another major criticism is the permissiveness of the US system in selling and sharing personal data for commercial purposes. In the US, user data is a valuable resource for companies, and many regulations do not provide sufficient restrictions to prevent unethical or abusive use.

- *The power of tech companies:* Some of the world's biggest tech companies, such as Google, Facebook (Meta), Amazon and Microsoft, are based in the US. These companies hold huge amounts of personal data and are often accused of invasive data collection and use practices. The lack of strict federal regulation in the US allows them to operate with greater freedom than would be allowed in the EU.

- *Government surveillance:* Another issue criticised is the access of government authorities to citizens' personal data in the name of national security. US legislation, such as the Patriot Act, allows the government to collect and use data for security purposes, which raises serious concerns about privacy violations. For example, the mass surveillance programme revealed by Edward Snowden has shed light on the extent of personal data collection by US government agencies.

US law compared with EU GDPR:

User access to data: the GDPR guarantees citizens' right of access to their data and gives them control over how it is used. In the US, this level of control and transparency is not widely available, with the exception of some state regulations, such as the CCPA.

Fines and penalties: the GDPR provides significant penalties for breaches, incentivising companies to comply with strict data protection rules. In the US, fines are rarer and often lower, except in cases of blatant violation of sector-specific laws.

User consent: the GDPR requires explicit and informed consent for the collection of personal data, whereas in the US, the consent policy is more relaxed and often tilted in favour of companies.

The lack of uniform federal legislation in the US creates a less predictable and less protective system for personal data privacy than the GDPR in the EU. Criticisms of insufficient data protection, combined with the influence of technology companies and concerns about government oversight, suggest an urgent

need for reform and a more rigorous federal legislative framework. However, despite these challenges, there is a growing movement in the US to improve data protection and bring standards in line with European ones.

5. Challenges in Implementing AI and Data Regulation

Artificial Intelligence (AI) has radically transformed the digital landscape and modern society, but applying data protection and privacy regulations in this context presents significant challenges. The complexity of AI technologies and their global nature have created major difficulties for regulators and organisations to comply.

Enforcement of privacy regulations in the AI era faces several obstacles that complicate the effectiveness and consistency of these laws, such as:

a. Opaque algorithms and regulatory difficulty. One of the biggest obstacles in regulatory enforcement is the opaque nature of AI algorithms. Many algorithms, especially those based on machine learning and neural networks, operate as ‘black boxes’. Roughly speaking, that an AI system is opaque means that it is difficult for users to know how it works, as well as to interpret its decisions at various levels and evaluate its behaviour against scientific and ethical norms.⁸

For example, algorithms can make decisions that have a significant impact on individuals, such as credit assessment, hiring or surveillance. The problem is that when automated decisions are challenged, companies are unable to provide clear explanations of how these decisions were made, in violation of the transparency and explainability requirements stipulated by modern regulations such as GDPR.

b. National and transnational borders. AI operates on a global scale, which makes enforcement of regulations extremely complicated, especially when collecting and transferring personal data between jurisdictions with different legal standards.

For example, in the European Union, the GDPR imposes strict rules for data protection, but other countries, such as the US, have a more relaxed approach and a fragment. This discrepancy creates a situation where a company operating globally has to comply with multiple conflicting regulations, which can be logistically difficult and costly.

For international data transfers, many companies use data infrastructure located in several countries. In this context, the transfer of personal data from the

⁸ Carlos Zednik (2021), "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence." *Philosophy & Technology* 34, no. 2 (2021): 265+. *Gale Academic One-File* (accessed July 20, 2025). <https://link.gale.com/apps/doc/A666288938/AONE?u=anon~35db595a&sid=googleScholar&xid=374b64c2>; Alessandro Facchini, Alberto Termine (2021), "Towards a taxonomy for the opacity of AI systems," in Vincent C. Müller (ed.), *Philosophy and Theory of Artificial Intelligence (PTAI)*, Ed. Cham, CH: Springer International Publishing, 2022, pp. 73–89.

EU to countries that do not offer an adequate level of protection becomes an issue. Despite international agreements such as the Privacy Shield (replaced by the Transatlantic Data Privacy Framework), there is a risk that citizens' personal data could be exposed to government surveillance or unauthorised uses.

c. Emerging technologies and delayed regulation. The rapid pace of AI development means that legislation often lags behind. Emerging technologies such as natural language processing, facial recognition, and autonomous vehicles present unique challenges that have not yet been fully addressed by existing regulations.

Current laws, such as GDPR, were written with a specific type of data processing in mind and do not fully cover the complexity of AI. For example, GDPR does not explicitly provide rules for how AI must be programmed or audited to prevent discrimination.

Regulators often find it difficult to react quickly to new challenges posed by AI due to lengthy and complex legislative processes. In addition, the lack of technical expertise of some regulators complicates the assessment of new technologies and their impact on privacy.

Although the implementation issues are complex, there are some directions and proposals that could contribute to better regulation of AI:

1. *Algorithmic transparency and auditability:* one suggested solution is the imposition of clear auditing standards for algorithms, which would allow verification of how AI processes data and makes decisions. This would require collaboration between authorities and technology experts to develop effective evaluation methods.

2. *Global regulatory standardisation:* The development of an international regulatory framework for AI could reduce differences across jurisdictions. Organisations such as the UN and the Council of Europe have already started discussing the creation of global principles for the ethical and responsible use of AI.

3. *Education and training of regulatory experts:* Regulators could benefit from better training and technology education to better understand and manage AI challenges.

The challenges of implementing AI and data regulation are complex and varied, involving algorithm transparency, international legal differences, and the rapid pace of technological innovation. In order to effectively protect citizens' privacy and maintain trust in AI technologies, there is a need for close collaboration between lawyers, engineers and regulators, as well as constant adaptation of regulations to new technological realities.

6. Algorithmic Bias and Discrimination

Algorithmic bias (or algorithmic bias) occurs when Artificial Intelligence algorithms make decisions that disproportionately disadvantage certain groups of

people, either because of the data used for training or because of the way these algorithms are designed. Because AI learns from large sets of data that may reflect existing inequalities and biases in society, there is a risk that these biases will be amplified, which can lead to significant discrimination in various domains.

AI has often led to discrimination, emphasising the importance of auditing and transparency of algorithms.

Examples of discrimination cases:

Criminal Risk Assessment Algorithm (COMPAS). The COMPAS algorithm has been used in several US states to assess a defendant's risk of recidivism and to help judges make bail or parole decisions.

An investigation by *ProPublica* in 2016 found that the algorithm was biased against people of colour. Although it was similarly accurate in predicting recidivism among different racial groups, COMPAS disproportionately classified people of colour as having a higher risk of committing future crimes, even when they did not reoffend. Instead, the algorithm tended to underestimate the risk for white people. This algorithmic bias raised major concerns about the fairness of the criminal justice system.⁹

The COMPAS case has emphasised the need for transparency and auditing of algorithms used in criminal justice systems, especially when they have a direct impact on people's lives. It also demonstrated the risk of using AI without understanding and correcting implicit biases in training data.

Hiring algorithms (Amazon). Amazon has developed an AI-based recruitment algorithm to evaluate candidates for technical positions. The algorithm was trained on CV data from the last 10 years of company employees.

The algorithm was found to discriminate against women, as the historical data on which it was trained reflected male dominance in technical positions. As a result, the AI learned to penalise CVs that contained words associated with women, such as 'women's chess club captain' or the names of women's universities. Even though these penalty criteria were not explicitly programmed, the algorithm absorbed biases from the training data.¹⁰

Amazon was forced to stop using the algorithm, but the case showed how easily discrimination can occur when biased data is used to train AI. Continuous auditing and assessing the fairness of algorithms are essential to prevent such negative effects.

Facial recognition and racial bias. Facial recognition technologies developed by companies such as IBM, Microsoft and Amazon have been tested for

⁹ Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin (2016), *How We Analysed the COMPAS Recidivism Algorithm*, <https://www.propublica.org/article/how-we-analyzed-the-comp-as-recidivism-algorithm>.

¹⁰ Maude Lavanchy (2018), *Amazon's sexist hiring algorithm could still be better than a human expecting algorithms to perform perfectly might be asking too much of ourselves*, The Conversation, Lausanne, Switzerland, <https://imd.widen.net/view/pdf/z7itobahi6/tc061-18-print.pdf>.

their accuracy in recognising the facial features of people from different ethnic groups.

Studies, such as the one conducted by the *MIT Media Lab*, showed that these algorithms had significantly higher error rates in correctly identifying people of colour, particularly women of colour. For example, a black woman had up to a 34% probability of being misclassified, while the error rate for white men was less than 1%. This bias can lead to serious results when technology is used by law enforcement agencies to identify suspects.¹¹

In response to these findings, some companies have suspended the sale of facial recognition technologies to police, emphasising the importance of ensuring the fairness of algorithms and testing them on diverse datasets.

Lending schemes and financial discrimination. Some financial technology companies are using AI to assess customer creditworthiness and decide whether to approve loans or credit cards. Algorithms analyse various data, including financial history, location, occupation and other behavioural variables.

One example is where lending algorithms gave lower scores to women compared to men, even when both groups had similar financial profiles. A high-profile case was that of Apple Card's lending programme, where several users reported that women were given lower credit limits than men, despite having comparable financial histories.¹²

This type of discrimination emphasises the need for regulation and oversight of the algorithms used for financial decisions. AI needs to be scrutinised to ensure that it does not introduce biases that affect access to financial resources.

Audit and transparency of algorithms are important for the following reasons:

1. Bias detection and correction: auditing algorithms may reveal biases that are not obvious at first glance. Continuous testing on diverse datasets can help to identify and eliminate bias.
2. Building trust: algorithmic transparency enables users and authorities to understand how automated decisions are made, thus helping to build trust in the use of AI. This is especially essential in sensitive sectors, such as justice, healthcare and finance.
3. Ethical and legal compliance: algorithms used by organisations must comply with data protection regulations and fairness principles. Without transparency and auditing, it is difficult to demonstrate that AI is acting in a fair and ethical way.

Algorithmic bias and discrimination caused by AI represent significant challenges that emphasise the urgent need to develop robust audit mechanisms and impose transparency requirements. Algorithms are not neutral; they are a

¹¹ <https://www.media.mit.edu/articles/study-finds-gender-and-skin-type-bias-in-commercial-artificial-intelligence-systems/>, accessed on 07.05.2025.

¹² <https://www.technologyreview.com/2019/11/11/131983/apple-card-is-being-investigated-over-claims-it-gives-women-lower-credit-limits/>, accessed on 07.05.2025.

product of the data they learn and the people who develop them. Without appropriate measures, AI risks perpetuating and even amplifying social inequalities, which requires an ethical and well-regulated approach.

7. Sources of Conflict Between Regulation and Innovation

In the age of Artificial Intelligence, strict data protection and privacy regulations play a crucial role in protecting users' rights and maintaining public trust in digital technologies. However, regulations can conflict with the rapid pace of technological innovation, creating a dilemma between user safety and technological progress. Let us analyse how these two issues interact and what sources of conflict arise.

a) Strict regulations can inhibit innovation. Stringent data protection regulations, such as GDPR in the European Union, impose complex requirements on companies developing AI technologies, which can lead to significant barriers to innovation.

We exemplify some ways in which these regulations can inhibit progress:

Compliance costs: Complying with strict regulations requires large investments in infrastructure and skilled staff to manage data privacy. For example, companies must hire data protection experts and implement complex security, auditing and reporting mechanisms. These requirements can be a heavy burden for startups and small companies that lack the resources to meet these standards. As a result, many startups may be discouraged from innovating in AI.

Slowing product development: Companies developing AI solutions need to conduct privacy impact assessments and implement preventive measures to minimise risks. These processes can delay new product launches and reduce the ability to compete in a dynamic global marketplace where rapid innovation is the key to success.

Limiting the use of data: Many regulations restrict how companies can collect and use users' personal data, which can hinder the development of advanced AI algorithms. AI depends on large amounts of data to learn and improve. When access to this data is restricted, AI innovation can suffer. For example, the GDPR's limitations on international data transfer can make it difficult to collect the diverse data needed to build efficient and fair AI systems.

b) Need for Regulation to Protect Users. Despite the impact on innovation, strict regulation is essential to protect users from the inherent risks of AI technologies. Some reasons for this need include:

Preventing abuse: AI has the ability to invade people's privacy in unprecedented ways by collecting and analysing personal data. Without clear regulations, companies could use this data in a way that jeopardises users' privacy and safety. Strict regulations ensure that data is processed transparently, fairly and only for well-defined purposes.

Reducing algorithmic discrimination: As discussed above, AI can reproduce or even amplify existing biases in the data it processes. Regulations require measures to prevent discrimination and ensure transparency of algorithms, thus protecting vulnerable groups from unfair or biased decisions.

Increased accountability: Without strict regulations, companies may avoid taking responsibility for errors or abuses of their algorithms. Clear rules, such as the right to explainability and the right to challenge automated decisions, force companies to be more transparent and accountable in their use of AI.

8. Relevant Cases of Conflict Between Regulation and Innovation

Healthcare industry and personal data: Companies developing AI technologies for the diagnosis and treatment of diseases rely on access to large sets of medical data to train and refine algorithms. However, GDPR and other medical privacy regulations limit access to patient data, slowing the pace of medical innovation. While these regulations protect patient privacy, they create a conflict with the urgent need to develop advanced AI solutions to save lives.

Autonomous vehicles and legal liability: Developers of autonomous vehicles face strict regulations on safety and legal liability. The algorithms that drive these vehicles need to be highly sophisticated and well tested before widespread deployment. However, strict regulations can delay testing on public roads and limit progress, although they are necessary to ensure public safety. This delicate balance between innovation and regulation continues to be a major challenge for the transport industry.

Financial technology (FinTech) and user data: FinTech companies are using AI to analyse financial behaviours and offer personalised credit solutions. However, data protection regulations sometimes prevent the efficient use of financial information, which limits AI's ability to personalise and improve services. Thus, innovation in FinTech can be hampered, even though these regulations are meant to protect users from financial abuse and privacy breaches.

Examples and the perspective of tech companies:

- *Google and Privacy Sandbox:* Google has announced initiatives such as Privacy Sandbox, to limit the use of third-party cookies and better protect user privacy¹³. But tech developers and advertisers have criticised the measure, arguing that it could inhibit innovation in the digital advertising industry and hurt the revenue streams of many online companies.

- *Facial recognition companies:* Some US cities, such as San Francisco, have banned the use of facial recognition technologies by government agencies because of privacy and civil rights risks. Companies that develop such technologies argue that such bans may limit innovation in public safety, though

¹³ <https://usercentrics.com/knowledge-hub/what-is-google-privacy-sandbox/>, accessed on 07.05.2025.

critics argue that the measures are necessary to prevent abuse.

While strict regulations may inhibit innovation in some areas, they are essential to protect users from potential AI abuse. One solution could be to adopt a flexible regulatory framework that encourages innovation while maintaining a high level of protection of personal data. Collaboration between regulators and the technology industry could also facilitate the development of solutions that respect both users' rights and the need for technological progress.

9. Conclusions

A global regulatory framework is becoming increasingly necessary as artificial intelligence advances and becomes a critical component in various economic and social sectors. The development and use of AI bring both significant opportunities and considerable risks, particularly in terms of data protection, ethics, and labour market impacts.

Regulating AI globally is important for several reasons:

Harmonisation of standards: as AI technology is used on a global scale, there is a risk that divergent regulations between countries could lead to legal issues, compliance difficulties and data protection vulnerabilities. A global framework could harmonise these regulations, ensuring effective protection for users regardless of region.

Cross-border risk management: AI can have effects that transcend national borders, such as information manipulation, cyber-attacks or influencing financial markets. A coordinated approach would help minimise these risks.

Data protection: Personal data is widely used by AI algorithms. A global framework could ensure clear rights for users and limits on data collection and processing.

Some key proposals for such a global framework include:

International co-operation and common standards:

- Creating an international platform for cooperation between governments, international organisations and private actors to establish common standards on AI and data protection.

- Building on existing initiatives, such as the European Union's General Data Protection Regulation (GDPR), which has set a precedent for the protection of personal data.

Algorithm transparency and accountability:

- A global requirement for companies developing AI to provide transparency about how their algorithms work and how personal data is used.

- Create independent audits to verify the impartiality and fairness of AI systems, thus preventing discrimination or systematic errors.

Ethical rules and user rights:

- Establish ethical principles such as respecting human dignity and ensuring that AI is used for the common good.

- The right of users to be informed when interacting with an AI system, as well as the option to refuse automation that significantly affects their lives.

Investment in education and workforce adaptation:

- International programmes to support the retraining of employees affected by automation to minimise the social and economic impact.
- Promote technology education and digital literacy to prepare new generations to operate with and around AI.

Cyber security and critical infrastructure:

- Implement cybersecurity policies to protect critical infrastructure from possible coordinated attacks involving AI.
- Information sharing between states to quickly and effectively manage potential threats.

Thoughtful global regulation could facilitate responsible innovation in AI, minimising risks and ensure benefits for society as a whole. The need for international coordination is evident to prevent a fragmented regulatory landscape that could disadvantage some states or groups of citizens.

Education and public awareness also play a key role in managing the impact of digital technologies, especially in the rapidly developing context of artificial intelligence and personal data protection. Informing citizens about their digital rights not only empowers them to protect their personal information, but also contributes to the empowerment of actors developing and using advanced technologies.

This requires the implementation of policies for education and awareness-raising, such as:

Educational programmes in schools and universities:

- Introduction of mandatory modules on digital rights and cybersecurity in school and university curricula.
- Organise practical workshops in schools to teach young people how to manage their digital footprint and how to recognise potential online threats.

National awareness campaigns:

- Governments and international organisations can initiate information campaigns through media, social media and public events to educate citizens about data protection.
- Develop partnerships with NGOs and technology companies to share relevant resources and information.

Easily accessible information platforms:

- Create online platforms where citizens can learn about their digital rights, how to protect their information, and how to report abuses or security breaches.
- Publish clear guides, translated into different languages, to ensure that information is accessible to the widest possible audience.

Workshops and courses for adults:

- Organising workshops for adults in local communities so that people of

all age groups learn how to navigate the digital environment safely.

- Free or subsidised courses to teach the public about data protection, such as using passwords correctly, recognising phishing attempts and securing personal devices.

So a well-informed population is less susceptible to manipulation, fraud and cyber attacks. Education increases digital resilience, protecting both individuals and society's critical infrastructure. When citizens are informed about their rights, they are more likely to actively participate in digital policy discussions and demand better protection and regulation. Understanding the ethical implications of technology and how personal data is used helps citizens to make more informed decisions and support the responsible use of AI. An educated public can also more effectively advocate for policies and regulations that promote data protection and ethics in the use of technology.

Bibliography

1. Acquisti, Alessandro & Jens Grossklags (2003), *Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior*, UC Berkeley 2nd Annual Workshop on "Economics and Information Security", available at: https://infosecon.net/workshop/downloads/2003/pdf/Final_session6_acquisti.pdf, accessed on 07.05.2025.
2. Baik, Jeeyun (Sophia) (2020), „Data privacy against innovation or against discrimination?: The Case of the California Consumer Privacy Act (CCPA)“, *Telematics and Informatics*, Vol. 52, 101431, <https://doi.org/10.1016/j.tele.2020.101431>.
3. Dienlin, Tobias & Philipp K. Masur, Sabine Trepte (2023). „A longitudinal analysis of the privacy paradox“. *New Media & Society*, 25(5), 1043-1064. <https://doi.org/10.1177/14614448211016316>.
4. Facchini, Alessandro & Alberto Termine (2021), "Towards a taxonomy for the opacity of AI systems," in Vincent C. Müller (ed.), *Philosophy and Theory of Artificial Intelligence (PTAI)*, Ed. Cham, CH: Springer International Publishing, 2022, pp. 73–89.
5. Larson, Jeff, Surya Mattu, Lauren Kirchner and Julia Angwin (2016), *How We Analysed the COMPAS Recidivism Algorithm*, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.
6. Lavanchy, Maude (2018), *Amazon's sexist hiring algorithm could still be better than a human expecting algorithms to perform perfectly might be asking too much of ourselves*, The Conversation, Lausanne, Switzerland, <https://imd.widen.net/view/pdf/z7itobahi6/tc061-18-print.pdf>.
7. McCarthy-Jones, Simon (2019) „The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century“, *Frontiers in Artificial Intelligence*, Vol. 2, <https://doi.org/10.3389/frai.2019.00019>.
8. Minkinen, Matti (2019). „Making the future by using the future: A study on influencing privacy protection rules through anticipatory storylines“. *New Media & Society*, 21(4), 984-1005. <https://doi.org/10.1177/1461444818817519>.
9. Peacock, Anne (2019). *Human rights and the digital divide*. London, Routledge,

- <https://doi.org/10.4324/9781351046794>.
10. Yeh, Chih-Liang (2018), „Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers”, *Telecommunications Policy*, Vol. 42, Issue 4, pp. 282–292, <https://doi.org/10.1016/j.telpol.2017.12.001>.
 11. Zednik, Carlos (2021), "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence." *Philosophy & Technology* 34, no. 2 (2021): 265+. *Gale Academic OneFile* (accessed July 20, 2025). <https://link.gale.com/apps/doc/A666288938/AONE?u=anon~35db595a&sid=googleScholar&xid=374b64c2>.

Artificial Intelligence Act and GDPR: Implications for AI Solution Developers and Users in Romania

Associate professor **Camelia Daciana STOIAN**¹

Professor **Dominic BUCERZAN**²

Lecturer **Radu Nicolae STOIAN**³

Assistant professor **Catalin Raul HALIC**⁴

Associate professor **Crina Anina BEJAN**⁵

Abstract

Artificial Intelligence (AI) poses a significant challenge for personal data protection legislation, substantially impacting the way Romanian companies develop and implement AI solutions, as well as affecting human rights. At the European level, the Artificial Intelligence Act (AIA)⁶ introduces a regulatory framework for the responsible use of AI, which must be harmonized with the General Data Protection Regulation (GDPR). In this context, Romania faces challenges regarding the compatibility of its national legislation with these European regulations, particularly concerning automated data processing, algorithmic transparency, user rights, and the impact of AI use in judicial and administrative systems. The study examines the extent to which Romanian legislation is prepared to accommodate the new requirements imposed by the AIA, highlighting legal risks and additional obligations for companies developing AI-based solutions. It also evaluates the potential consequences for the Romanian technology market, including impacts on AI-focused startups and institutions utilizing artificial intelligence technologies in their operational processes. The study's conclusions emphasize the need for a proactive and integrated approach to ensure compliance with European standards while simultaneously protecting technological innovation and user rights.

Keywords: artificial intelligence, data protection, Artificial Intelligence Act, GDPR, Romanian legislation, AI regulation.

JEL Classification: K20, K23, K24

¹ Camelia Daciana Stoian - Faculty of Humanities and Social Sciences, "Aurel Vlaicu" University of Arad, Romania, ORCID: 0000-0003-2776-6244, av.stoiancameliadaciana@yahoo.com.

² Dominic Bucerzan - Faculty of Exact Sciences, "Aurel Vlaicu" University of Arad, Romania, ORCID: 0000-0002-9260-9387, dominic@bbcomputer.ro.

³ Radu Nicolae Stoian - Faculty of Law, "Vasile Goldiș" University of Arad, Romania, radustoian73@gmail.com.

⁴ Catalin Raul Halic - Faculty of Exact Sciences, "Aurel Vlaicu" University of Arad, Romania, ORCID: 0009-0002-0459-2464, haliccatalin@gmail.com.

⁵ Crina Anina Bejan - Faculty of Exact Sciences, "Aurel Vlaicu" University of Arad, Romania, ORCID: 0000-0003-2868-2376, ratiu_anina@yahoo.com.

⁶ Regulation of the European Parliament and of the Council of the European Union, No. 1689 of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Regulation).

Please cite this article as:

Stoian, Camelia Daciana, Dominic Bucerzan, Radu Nicolae Stoian, Catalin Raul Halic & Crina Anina Bejan, „Artificial Intelligence Act and GDPR: Implications for AI Solution Developers and Users in Romania”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 162-177.

1. Introduction

In recent years, artificial intelligence (AI) has rapidly evolved from a niche research field into a transformative force across multiple sectors, including healthcare, finance, education, and public administration⁷. Its capacity to process vast quantities of data, identify patterns, and make autonomous decisions introduces not only substantial opportunities but also serious legal and ethical concerns. Among the most pressing is the challenge of ensuring that AI systems respect fundamental rights, particularly the right to personal data protection, as enshrined in both national and European legal frameworks⁸. Although a growing body of academic work addresses the normative and theoretical implications of AI regulation, the current literature lacks comprehensive empirical studies that explore its impact across diverse organisational contexts⁹ — particularly in smaller markets such as Romania. This gap underscores the importance of examining how emerging legal frameworks interact with real-world technological development.

At the European level, two major legal instruments shape the governance of artificial intelligence and data protection: the General Data Protection Regulation (GDPR)¹⁰, which has been in force since 2018, and the newly adopted Artificial Intelligence Act (AIA)¹¹, the first comprehensive legal framework for AI

⁷ Yiming Yuan, Yongming Sun and Hangyu Chen. 2024. “Does Artificial Intelligence Affect Firms’ Inner Wage Gap?” *Applied Economics* 57 (19): 2365–71. doi: 10.1080/00036846.2024.2324090.

⁸ Abdallah Q. Bataineh, Alaa S. Mushtaha, Ibrahim A. Abu-AlSondos, Saeed Hameed Aldulaimi, Marwan Abdeldayem. 2024. “Ethical & Legal Concerns of Artificial Intelligence in the Healthcare Sector,” *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, Manama, Bahrain, pp. 491-495, doi: 10.1109/ICETISIS61505.2024.10459438.

⁹ João Pedro Quintais 2025. “Generative AI, copyright and the AI Act.” *Computer Law & Security Review*, vol. 56: 106107, <https://doi.org/10.1016/j.clsr.2025.106107>.

¹⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

¹¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024.

within the European Union. While the GDPR provides strong protections for personal data, including limitations on profiling and automated decision-making, the AIA introduces a risk-based approach to AI systems, imposing specific requirements on applications considered “high-risk”. The interplay between these two regulations reflects the EU’s broader commitment to promoting trustworthy AI that aligns with democratic values and fundamental rights. However, the simultaneous applicability of both instruments also creates legal complexity — particularly for organisations tasked with ensuring compliance in practice¹².

For Romania, the implementation of these regulatory frameworks presents unique challenges. As an EU member state with a rapidly developing tech ecosystem, Romania must align its national legislation and institutional practices with the obligations introduced by both the GDPR and the AIA. However, the current legal infrastructure lacks specific provisions addressing issues such as algorithmic transparency, automated decision-making, or discrimination resulting from AI systems. This regulatory gap raises concerns about both compliance and the protection of individual rights¹³.

In addition to the legal dimension, the regulation of AI in Romania has direct implications for the country’s economic ecosystem. Romania is home to both companies that develop AI solutions — such as automation, natural language processing, or behavioural authentication technologies—and organisations across sectors that rely on AI tools in their operations¹⁴. The integration of the GDPR and AIA into national practice is therefore not merely a matter of legal compliance, but one that will influence innovation, competitiveness, and the ability of local firms to scale within the EU digital market^{15,16}.

Romania, as both an emerging market and an EU member state, finds itself at a crossroads between significant technological potential and persistent legal and institutional challenges. In this context, the purpose of this article is to assess the readiness of Romanian legislation and institutional frameworks to accommodate the requirements of the Artificial Intelligence Act in conjunction with

¹² Lena Enqvist. 2024. "Rule-based versus AI-driven benefits allocation: GDPR and AIA legal implications and challenges for automation in public social security administration." *Information & Communications Technology Law* vol. 33, no. 2: 222-246, doi: 10.1080/13600834.2024.2349835.

¹³ Anca Parmena Olimid, Catalina Maria Georgescu, and Daniel Alin Olimid. 2024. "Legal Analysis of EU Artificial Intelligence Act (2024): Insights from Personal Data Governance and Health Policy." *Access to Justice in Eastern Europe* 7(4): 120-42 <<https://doi.org/10.33327/AJEE-18-7.4-a000103>>.

¹⁴ Daniel Castro and Michael McLaughlin, “Who Is Winning the AI Race: China, the EU, or the United States?” Center for Data Innovation, January 2021, <https://datainnovation.org/2021/01/who-is-winning-the-ai-race-china-the-eu-or-the-united-states-2021-update/>.

¹⁵ Chambers and Partners. (2024). *Artificial Intelligence 2024 – Romania: Law & Practice Guide*. Available at: <https://practiceguides.chambers.com/practice-guides/artificial-intelligence-2024/romania> [Accessed 21 Mar. 2025].

¹⁶ Nick Wallace and Daniel Castro (2018). *The Impact of the EU’s New Data Protection Regulation on AI*. Information Technology and Innovation Foundation (ITIF). Available at: <https://itif.org/publications/2018/03/26/impact-eu-new-data-protection-regulation-ai> [Accessed 21 Mar. 2025].

the GDPR. The analysis also aims to explore the broader implications of this alignment for AI developers, users, and regulators, highlighting areas where proactive adaptation is essential.

2. The European Legal Framework: AIA and GDPR

The European Union has positioned itself as a global leader in regulating the digital environment, with the General Data Protection Regulation (GDPR) and the Artificial Intelligence Act (AIA) as two cornerstone instruments. While the GDPR focuses on safeguarding personal data and ensuring individual control over data processing, the AIA introduces a framework for the ethical and safe development, deployment, and use of AI technologies. These instruments are designed to work in tandem, reinforcing the EU's commitment to human-centric, trustworthy AI.

The Artificial Intelligence Act introduces a tiered risk classification model, dividing AI systems into minimal, limited, high, and unacceptable risk categories. The classification is based on the intended use of the AI system, its potential to affect fundamental rights, and the degree of autonomy involved. High-risk systems are typically those used in sensitive contexts such as biometric identification, access to education or employment, healthcare, and legal decision-making. Once designated as high-risk, these systems are subject to a set of mandatory compliance obligations that go beyond general ethical recommendations, forming binding legal standards.

Real-life examples of such risks include AI algorithms used in hiring platforms, which may inadvertently exclude candidates based on biased training data, as alleged in *Mobley v. Workday Inc.*¹⁷ and addressed in *EEOC v. iTutor-Group*¹⁸. In credit scoring, “black-box” models have drawn regulatory scrutiny from the U.S. Consumer Financial Protection Bureau (CFPB) for failing to provide explainable justifications for denied loans¹⁹. In the housing sector, *Louis v. SafeRent Solutions* revealed how tenant screening algorithms could systemically disadvantage applicants from minority backgrounds²⁰. These examples highlight the necessity for strong oversight and legal accountability in high-risk AI domains.

¹⁷ *Mobley v. Workday Inc.*, Case No. 23-cv-770 (N.D. Cal. 2023). Reuters report: <https://www.reuters.com/legal/litigation/workday-must-face-novel-bias-lawsuit-over-ai-screening-software-2024-07-15> [Accessed 16 Mar. 2025].

¹⁸ *EEOC v. iTutorGroup*, U.S. Equal Employment Opportunity Commission settlement (2023). ABA summary: https://www.americanbar.org/groups/business_law/resources/business-law-today/2024-april/navigating-ai-employment-bias-maze [Accessed 17 Mar. 2025].

¹⁹ CFPB Guidance on Credit Algorithms (2022). Consumer Finance Protection Bureau: <https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms> [Accessed 18 Mar. 2025].

²⁰ *Louis v. SafeRent Solutions LLC*, Case No. 1:22-cv-10760 (D. Mass. 2023). AP coverage: <https://apnews.com/article/1bc785c24a1b88bd425a8fa367ab2b23> [Accessed 18 Mar. 2025].

Effective data governance is a cornerstone of the Artificial Intelligence Act, particularly for high-risk AI systems. Developers are required to ensure that datasets used in training, validation, and testing are relevant, representative, and free from errors or distortions that could lead to biased outcomes. This is essential for preventing discriminatory or harmful outputs, especially in areas involving sensitive personal data. The AIA further mandates documentation of data provenance, justification for data collection methods, and traceability throughout the AI system's lifecycle.

The lack of proper data governance has already produced notable legal and regulatory consequences. In the case of *State v. Clearview AI*, multiple European data protection authorities fined and banned the facial recognition company for harvesting billions of images without consent — highlighting the importance of lawful and proportionate data collection practices^{21, 22}. Similarly, the *Netherlands SyRI* case invalidated a government-run risk prediction system for violating privacy rights due to opaque data use and lack of transparency²³. These examples reveal how flawed or unregulated data governance not only erodes public trust but also contravenes fundamental rights, placing both developers and users of AI systems at legal risk.

Transparency is a fundamental requirement for high-risk AI systems under the Artificial Intelligence Act. Developers must design systems that are not only technically robust but also capable of offering meaningful explanations of how decisions are made. This includes informing users that they are interacting with an AI system, clarifying the logic behind automated decisions, and enabling scrutiny by regulators and affected individuals. Explainability is particularly important when decisions significantly affect individuals' rights, such as access to credit, employment, or public services.

Legal disputes have demonstrated the dangers of opaque AI systems. In *Burdick v. Employment Development Department (EDD)*, California residents sued the state for relying on a flawed algorithm that wrongly denied unemployment benefits without meaningful explanation or recourse^{24, 25}. The court ruled that the system violated due process rights. Similarly, in the UK, the A-Level

²¹ Clearview AI cases brought by data protection authorities across the EU (2021–2023). Example: France CNIL decision (2022): <https://www.cnil.fr/en/clearview-ai-ordered-stop-reuse-facial-recognition-data-and-delete-data> [Accessed 18 Mar. 2025].

²² UK ICO enforcement: <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-clearview-ai-inc-over-breach-of-uk-data-protection-laws> [Accessed 18 Mar. 2025].

²³ Netherlands District Court of The Hague, ECLI:NL:RBDHA:2020:865 (SyRI case). Summary via Human Rights Watch: <https://www.hrw.org/news/2020/02/06/dutch-court-halts-dystopian-surveillance-system> [Accessed 18 Mar. 2025].

²⁴ *Burdick v. EDD*, U.S. District Court, Northern District of California, Case No. 3:21-cv-02808. News summary: <https://www.reuters.com/legal/government/california-sued-over-flawed-algorithm-used-deny-jobless-benefits-2021-04-19> [Accessed 18 Mar. 2025].

²⁵ Case summary via EFF: <https://www.eff.org/cases/burdick-v-california-edd> [Accessed 18 Mar. 2025].

grading scandal of 2020 — where an algorithm downgraded thousands of students' grades based on opaque criteria — sparked widespread public backlash and led to the abandonment of the model^{26,27}. These cases underscore the principle that algorithmic decision-making must be auditable, intelligible, and subject to human review when individual rights are at stake.

The Artificial Intelligence Act requires that high-risk AI systems include safeguards to ensure effective human oversight. This principle is based on the idea that human operators must remain meaningfully involved in decision-making processes, especially where the outcomes affect individuals' rights or safety. Oversight may involve the ability to interpret and contest AI outputs, intervene before harm occurs, or deactivate systems in real time. The AIA also mandates continuous risk management procedures, including the identification, assessment, and mitigation of foreseeable risks, as well as post-market monitoring and incident reporting.

Failure to implement such mechanisms has resulted in tangible harm. In Michigan Unemployment Insurance Agency (UIA), over 40,000 individuals were wrongly accused of fraud due to an automated system with no human oversight or appeal mechanism^{28,29}. The state was later required to issue mass reimbursements and conduct human reviews of past decisions. Similarly, in Australia's infamous "Robodebt" case, a government-run AI system used flawed income-averaging algorithms to issue unlawful debt notices to welfare recipients without proper human verification^{30,31}. A Royal Commission found systemic failures in governance, and the government ultimately repaid over AU\$1.7 billion to affected citizens. These cases reinforce the necessity of embedding human judgement and accountability into the design and deployment of high-risk AI.

Although the Artificial Intelligence Act introduces AI-specific rules, the General Data Protection Regulation (GDPR) remains a cornerstone of data protection within the EU and is highly relevant to the development and deployment of AI systems. Key GDPR principles — such as lawfulness, fairness, transparency, and purpose limitation — directly affect how AI systems may collect and

²⁶ UK A-Level Algorithm Scandal (2020). Coverage by BBC News: <https://www.bbc.com/news/education-53805> [Accessed 18 Mar. 2025].

²⁷ Analysis from The Guardian: <https://www.theguardian.com/education/2020/aug/17/algorithm-that-downgraded-a-level-results-must-never-be-used-again> 105 [Accessed 18 Mar. 2025].

²⁸ Michigan UIA scandal involving automated fraud detection (2013–2020). News summary: <https://www.freep.com/story/news/local/michigan/2020/08/10/michigan-jobless-agency-fraudulent-claims/3335021001> [Accessed 16 Mar. 2025].

²⁹ Legal coverage: https://www.wnem.com/news/michigan-to-reimburse-residents-wrongly-flagged-by-ai-fraud-system/article_4e8e18fa-2c8a-11ed-84de-0b5be66b2f28.html [Accessed 16 Mar. 2025].

³⁰ Australia Robodebt Royal Commission (2023). ABC News coverage: <https://www.abc.net.au/news/2023-07-07/robodebt-royal-commission-final-report-released/102567034> [Accessed 16 Mar. 2025].

³¹ Royal Commission Report (official): <https://robodebt.royalcommission.gov.au/publications/final-report> [Accessed 16 Mar. 2025].

process personal data. Particularly important are Article 22 GDPR, which grants individuals the right not to be subject to decisions based solely on automated processing, and Article 7, which governs conditions for valid consent. These provisions create clear legal boundaries for profiling, behavioural prediction, and algorithmic decision-making, requiring developers to incorporate safeguards such as human intervention, explanation, and contestability mechanisms.

Despite their shared goal of protecting fundamental rights, the AIA and GDPR differ in scope, structure, and enforcement mechanisms, occasionally leading to areas of overlap or regulatory tension. While the GDPR focuses on how personal data is processed, the AIA regulates the function and risk of the AI system as a whole, including those that do not necessarily involve personal data. However, in high-risk AI systems that due process personal data — such as biometric identification or credit scoring — the two instruments converge. One tension arises in relation to explainability: GDPR's transparency obligations require that individuals understand how decisions are made, while many AI systems operate as “black boxes” that defy easy interpretation. Additionally, ambiguity remains regarding how the two frameworks interact procedurally — for example, whether a system's AIA compliance can be interpreted as sufficient proof of GDPR compliance, or whether dual assessments are required. These uncertainties highlight the need for harmonised guidance and enforcement practices, particularly at the national level.

3. Romanian National Context: Legal Readiness and Gaps

Romania, as an EU member state, is directly subject to the provisions of both the GDPR and the forthcoming Artificial Intelligence Act. While the GDPR has been transposed and implemented through national legislation — particularly Law No. 190/2018, which provides national derogations and clarifications — the country currently lacks any dedicated legal framework for AI regulation. As of early 2025, there are no specific national laws governing algorithmic decision-making, transparency of AI systems, or the mitigation of algorithmic bias. In this context, the entry into force of the AIA presents both a legal and institutional challenge for Romania.

Despite Romania's alignment with the GDPR, the national legal framework remains silent on key aspects of AI governance, particularly in relation to algorithmic decision-making, transparency, and bias mitigation. Currently, there are no binding national provisions that define how automated decision systems should be audited, how their logic must be disclosed, or how discriminatory outcomes should be identified and prevented. The absence of such regulations creates a regulatory vacuum, especially in high-impact sectors like employment, credit, and public administration, where AI is already being deployed. Without legal clarity, Romanian companies and public institutions risk either under-regu-

lating, thereby infringing fundamental rights, or over-complying, which may stifle innovation due to legal uncertainty.

Romanian companies engaged in the development of artificial intelligence technologies face substantial uncertainty due to the absence of a dedicated national legal framework governing algorithmic transparency, accountability, and bias mitigation. In the current context, these entities must rely primarily on the GDPR and anticipate the future applicability of the Artificial Intelligence Act, yet they lack specific national guidance tailored to AI-specific compliance. This creates ambiguity regarding lawful data processing, model auditing obligations, and explainability standards. Consequently, many AI developers may adopt a risk-averse posture, slowing innovation and investment. For instance, Romanian startups such as TypingDNA, which builds AI-based behavioural authentication, and Druid AI, which develops conversational AI systems, operate in a regulatory vacuum with limited domestic support for legal risk management³². These companies must navigate legal uncertainty on their own or through external EU guidance, which adds operational complexity and potential compliance costs.

Organisations that integrate AI systems into their operations — especially in sectors like finance, recruitment, and e-commerce — also encounter regulatory and reputational risks due to the lack of national standards. AI adoption in Romanian businesses is steadily increasing, yet the absence of rules on algorithmic decision-making or profiling opens the door to inconsistent practices. For example, financial institutions experimenting with AI-driven credit scoring or risk assessment tools often do so without clear guidance on transparency or user rights. A 2023 report noted growing consumer complaints related to automated loan refusals and opaque decision-making in digital banking services in Romania³³. Without clear mechanisms for auditability and user recourse, such practices risk violating Articles 13–15 and 22 of the GDPR and may undermine public trust in AI-based services. In the absence of regulatory certainty, businesses are left to define their own compliance thresholds — an approach that may result in uneven protection of fundamental rights and reputational exposure.

Beyond individual companies, the regulatory vacuum has wider implications for Romania's economic positioning. In a highly competitive regional tech landscape, the lack of legal clarity in AI governance can act as a deterrent to both

³² TypingDNA develops AI-based typing biometrics used for behavioural authentication in security systems. See: <https://www.typingdna.com>. Druid AI builds conversational AI and NLP solutions for enterprises and raised €14.2 million in Series A funding to support global expansion. See: EU Startups, “Druid raises €14.2M to scale AI-driven chatbots,” (2022), available at: <https://www.eu-startups.com/2022/05/bucharest-based-druid-snaps-up-e14-2-million-for-its-innovative-ai-driven-chatbots-and-is-set-to-soar> [Accessed 16 Mar. 2025].

³³ Romanian Financial Supervisory Authority, Consumer Protection Division Reports (2023), summary data on digital finance complaints. See also public discussions in: HotNews.ro, “Credit digital refuzat automat? Lipsa de transparență la bănci poate atrage sancțiuni,” (May 2023), available at: https://economic.hotnews.ro/stiri-finante_banci-26258435-credite-digitale-refuzate-automat-lipsa-transparenței-poate-atras-sanctiuni.htm [Accessed 16 Mar. 2025].

domestic innovation and foreign direct investment. Investors and multinational partners typically require predictable and stable legal environments — particularly in emerging technology sectors. In response, Romanian authorities adopted the National Strategy on Artificial Intelligence for 2024–2027, aiming to harmonise domestic policy with EU digital objectives, including the implementation of the AI Act³⁴. The strategy highlights key focus areas such as digital public services, education, cybersecurity, and responsible AI development. However, as of early 2025, the strategy remains largely programmatic and lacks concrete legislative instruments or enforcement mechanisms. A proactive legal and institutional framework will be essential not only to attract investment but also to ensure ethical, lawful, and economically sustainable AI integration.

The institutional capacity to enforce data protection and future AI regulation in Romania remains limited. The National Authority for the Supervision of Personal Data Processing (ANSPDCP) is the primary body responsible for GDPR enforcement, but it has so far played a relatively modest role in the emerging debate around algorithmic accountability and AI oversight. Its enforcement actions have focused primarily on traditional data breaches, with limited public engagement or guidance regarding automated decision-making or profiling under Article 22 GDPR³⁵. The judiciary has also faced challenges in addressing complex data-driven cases, due to limited technical expertise and the novelty of AI-related disputes. In the public sector, algorithmic tools are being introduced (e.g., in tax administration or digital public services), yet no unified framework or oversight mechanism exists to evaluate their legality or impact. This institutional lag poses risks not only for rights protection but also for effective implementation of the AI Act once it becomes fully applicable.

In addition to limited institutional readiness, Romania faces challenges stemming from regulatory fragmentation and legal ambiguity. While several digital strategies and policy frameworks exist — such as the National AI Strategy and various e-Governance initiatives — these remain largely aspirational and are not supported by enforceable legal instruments. Moreover, the interaction between horizontal legal norms (such as GDPR) and emerging sector-specific policies (e.g., in finance or health) has not been clearly articulated in legislation or practice. This lack of coherence creates uncertainty for private and public actors alike, as they struggle to interpret how existing rules apply to AI systems in the absence of case law, regulatory guidance, or coordinated enforcement. As Romania prepares to align with the Artificial Intelligence Act, addressing these legal and institutional inconsistencies will be essential to avoid fragmented implemen-

³⁴ U.S. Department of Commerce, “Romania – Digital Economy: Country Commercial Guide,” (2024), available at: <https://www.trade.gov/country-commercial-guides/romania-digital-economy> [Accessed 16 Mar. 2025].

³⁵ ANSPDCP – Annual Activity Reports (2019–2023). Available at: <https://www.dataprotection.ro/?page=Raportare&lang=en>.

tation and to ensure both innovation and fundamental rights are adequately protected.

4. AI Use in the Romanian Tech Ecosystem

In the last decade, Romania has emerged as a regional hub for technology and innovation, with a growing number of startups and scale-ups developing AI-driven solutions. Several Romanian-founded companies have gained international visibility through the integration of artificial intelligence into software products. Notably, UiPath, originally founded in Bucharest, became a global leader in robotic process automation (RPA), incorporating AI to enhance document understanding, task mining, and decision-making processes. Other firms, such as Druid AI, which develops conversational AI platforms for enterprise clients, and TypingDNA, known for behavioural biometrics and continuous authentication, exemplify the innovative applications of AI originating from the Romanian ecosystem³⁶. These companies operate within or adjacent to the high-risk AI categories defined by the Artificial Intelligence Act, particularly in areas like workplace automation, identity verification, and customer interaction.

Romanian AI developers are active across a range of sectors, reflecting the country's growing integration into the European and global digital economy. In the enterprise automation space, UiPath remains the most prominent example, achieving “unicorn” status in 2018 and later listing on the New York Stock Exchange in 2021 — an achievement that brought international attention to the Romanian tech ecosystem. In the natural language processing and customer experience domains, Druid AI has expanded rapidly, securing €14.2 million in Series A funding in 2022 to support international expansion and product scaling³⁷. In the field of cybersecurity and behavioural analytics, TypingDNA offers authentication tools based on AI-powered keystroke dynamics, with applications in fintech, education, and secure enterprise systems. Other emerging firms, such as MorphL (acquired by Algolia in 2021), applied AI to personalise user experiences in e-commerce environments³⁸.

Funding for AI startups in Romania comes from a mix of EU-backed programmes, venture capital, and local accelerators such as Techcelerator, which supports early-stage AI ventures with seed funding and regulatory mentoring. While access to funding has improved, Romanian AI developers still face structural constraints related to limited legal infrastructure, underdeveloped public-

³⁶ UiPath: <https://www.uipath.com>, Druid AI: <https://www.druidai.com>, TypingDNA: <https://www.typingdna.com>.

³⁷ “Druid raises €14.2M to scale AI-driven chatbots,” EU Startups (2022). Available at: <https://www.eu-startups.com/2022/05/bucharest-based-druid-snaps-up-e14-2-million-for-its-innovative-ai-driven-chatbots-and-is-set-to-soar>.

³⁸ “Algolia acquires MorphL to personalize AI-powered search,” TechCrunch (2021). Available at: <https://techcrunch.com/2021/01/26/algolia-acquires-morphl/>.

private partnerships, and a domestic market that remains risk-averse in adopting emerging technologies. These factors limit the scalability and long-term competitiveness of AI enterprises unless accompanied by targeted regulatory and institutional support.

The AI systems developed by Romanian companies are likely to fall under the “high-risk” category as defined by the Artificial Intelligence Act. These include applications used in areas such as biometric identification, access to financial services, employment, and customer profiling. For example, TypingDNA’s behavioural biometrics, used for continuous authentication in finance and education, involves sensitive personal data and could be classified as high-risk under the AIA due to its potential impact on access to essential services and data protection rights³⁹. Similarly, Druid AI’s conversational platforms —when integrated into hiring platforms or health-related services— may be subject to stricter obligations depending on their deployment context. Even UiPath’s process automation tools, while general-purpose in nature, may fall under the AIA’s scope if used in judicial or public administrative settings where automated decision-making has legal consequences.

The AIA’s risk-based framework places considerable responsibility on developers to assess the intended use of their products and apply appropriate compliance mechanisms. This includes risk management systems, detailed documentation, human oversight protocols, and post-market monitoring —requirements that may place a disproportionate burden on small and medium-sized enterprises (SMEs), which make up the majority of Romania’s AI innovation landscape. Without national implementation guidelines or regulatory support structures, Romanian developers risk falling behind in both compliance and competitiveness, particularly when seeking to scale within the EU market.

AI adoption in Romania is no longer limited to developers — large companies and public institutions have also begun integrating AI technologies into their operational workflows. In the banking sector, institutions such as Banca Transilvania and BRD – Groupe Société Générale have implemented AI-driven tools for fraud detection, customer service chatbots, and credit scoring, aiming to improve efficiency and user experience⁴⁰. Similarly, in the telecommunications industry, companies like Orange Romania and Vodafone have deployed AI for network optimisation, predictive maintenance, and customer engagement, including through virtual assistants and intelligent routing systems. These use cases illustrate Romania’s growing reliance on AI not just as a back-end optimisation

³⁹ For an overview of AIA’s “high-risk” categories, see: European Commission, “Proposal for a Regulation laying down harmonised rules on artificial intelligence,” (COM/2021/206 final). Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.

⁴⁰ See Banca Transilvania AI chatbot “Raul,” launched for customer support services. Coverage: ZF Tech, “Banca Transilvania lansează Raul, asistent virtual bazat pe AI,” (2021), available at: <https://www.zf.ro/business-hi-tech/banca-transilvania-lanseaza-raul-un-asistent-virtual-care-utilizeaza-19913929>. See also: Orange Romania AI-driven customer services. Company new room: <https://www.orange.ro/newsroom>.

tool, but as a direct interface between service providers and consumers.

In the public sector, digitalisation strategies have included elements of AI integration, especially in areas such as tax administration (e.g., automated document processing at ANAF, the National Tax Administration Agency), health system logistics, and smart city initiatives.

However, the broader implementation of the GDPR and the forthcoming AIA introduces significant compliance obligations that can reshape business models and affect the cost-benefit calculus of adopting AI. Under these regulations, companies must assess whether their systems fall under high-risk classifications, ensure lawful data processing, and implement transparency, human oversight, and risk mitigation mechanisms. These requirements may increase operational costs, especially for smaller firms, while also raising concerns over legal liability and reputational risks. As a result, some companies may limit or delay AI deployment, particularly in sensitive areas such as finance or HR, where the stakes of non-compliance are higher. Balancing innovation with regulatory risk has thus become a key strategic consideration in Romania's evolving digital economy.

The integration of AI into the Romanian public sector presents both opportunities and substantial risks. Institutions such as the National Agency for Fiscal Administration (ANAF) and various municipal governments have begun exploring AI applications for document automation, service delivery optimisation, and predictive analytics. There are also discussions around AI-supported systems for case management in courts, digital legal research, and even resource allocation in public employment processes. However, in the absence of robust legal and ethical frameworks, the deployment of such systems risks violating principles of due process, non-discrimination, and administrative transparency.

Compliance with the Artificial Intelligence Act and GDPR will require public authorities to conduct fundamental rights impact assessments, ensure algorithmic transparency, and provide mechanisms for human oversight and contestability. These obligations may necessitate the creation of internal compliance units, staff retraining, and collaboration with external regulators—demands that many Romanian public institutions are currently ill-equipped to meet. Without proactive institutional adaptation, the use of AI in governance may exacerbate systemic inefficiencies or deepen existing social inequalities, rather than resolving them. As such, the adoption of AI in the public sector must be accompanied by a clear strategy for legal compliance, ethical alignment, and accountability.

5. Challenges and Strategic Directions for AI Governance in Romania

One of the most pressing challenges Romania faces in the context of AI

governance is the lack of established mechanisms to ensure algorithmic transparency and explainability⁴¹. While the GDPR mandates user information rights and safeguards against fully automated decision-making (Article 22), these provisions are rarely enforced in practice. Moreover, the Artificial Intelligence Act introduces further requirements — such as risk classification, logging, and documentation — that public and private actors in Romania are largely unprepared to meet. Many existing AI systems, especially those procured or developed without a legal compliance framework, function as “black boxes,” where decision logic is opaque even to their implementers. This undermines accountability, particularly in sectors like finance, employment, and public services where algorithmic decisions can significantly affect individual rights.

Another critical concern is the risk of algorithmic discrimination⁴², particularly when AI systems are trained on biased or non-representative data. In Romania, this issue is amplified by the absence of formal auditing requirements or standardised evaluation procedures for bias detection. High-risk domains — such as credit scoring, recruitment, and welfare allocation — are especially vulnerable to unjustified disparities in outcomes. For instance, AI models used for pre-screening job applicants may inadvertently disadvantage certain demographic groups, while automated credit assessments could embed historical inequalities due to reliance on legacy datasets. Without dedicated national guidance, these risks remain difficult to identify and even harder to remedy, especially for smaller entities lacking internal legal or ethical oversight capacity.

Beyond technical and legal challenges, Romania’s public institutions are limited in their capacity to implement, supervise, and enforce AI-related obligations⁴³. Regulatory bodies such as ANSPDCP currently lack the specialised personnel and technical infrastructure to audit AI systems or issue sector-specific guidance. The judiciary, too, faces obstacles in adjudicating AI-related cases, which often require multidisciplinary expertise that is not yet integrated into judicial training. Similarly, most public institutions deploying AI do so without dedicated ethics committees, risk impact protocols, or transparent procurement rules. This institutional inertia could delay the effective enforcement of both GDPR and the AIA, undermining Romania’s compliance with EU digital policy goals.

To address these challenges, Romania must adopt a proactive and coher-

⁴¹ Polat Goktas. 2024. “Ethics, Transparency, and Explainability in Generative Ai Decision-Making Systems: A Comprehensive Bibliometric Study.” *Journal of Decision Systems*, October, 1–29. doi: 10.1080/12460125.2024.2410042.

⁴² Xukang Wang, Ying Cheng Wu, Xueliang Ji, Hongpeng Fu. 2024. "Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices." *Frontiers in Artificial Intelligence*, vol. 7: 1320277, <https://doi.org/10.3389/frai.2024.1320277>.

⁴³ Ahmed Oudah Mohammed Al-Dulaimi, Mohammed Abd-Al Wahab Mohammed. 2025 „Legal responsibility for errors caused by artificial intelligence (AI) in the public sector”. *International Journal of Law and Management*, <https://doi.org/10.1108/IJLMA-08-2024-0295>.

ent strategy for AI governance. First, national legislation should explicitly integrate the obligations set forth in the AIA and clarify their relationship with existing data protection laws. Second, sector-specific regulatory guidelines should be developed — particularly in high-risk areas such as finance, education, and public administration — outlining best practices for transparency, data governance, and human oversight. Third, public investment should focus on institutional capacity-building, including the creation of expert units within regulators and the judiciary, as well as funding for algorithmic auditing infrastructure.

Finally, legal reform must be accompanied by economic and educational support mechanisms. This includes establishing regulatory sandboxes for AI innovation, where startups and SMEs can test high-risk systems under regulatory supervision; offering compliance toolkits for companies with limited in-house legal capacity; and integrating AI ethics and regulation into academic and professional training programs. These measures will help ensure that Romania's AI ecosystem remains competitive and responsible. Striking the right balance between innovation and rights protection is essential — not only for legal compliance with the AIA and GDPR, but for the long-term legitimacy and public acceptance of AI technologies in Romanian society.

6. Conclusions

Romania enters a new regulatory era defined by the interplay between the General Data Protection Regulation and the forthcoming Artificial Intelligence Act. In this environment it faces both significant challenges and valuable opportunities. While the country has demonstrated technological potential through its emerging AI startup ecosystem and growing digital infrastructure, its legal and institutional frameworks remain underdeveloped in key areas such as algorithmic transparency, bias mitigation, and risk accountability.

This paper has argued that the lack of national regulation specific to AI — and the limited institutional capacity to interpret and enforce EU-level standards — poses legal, economic, and societal risks. At the same time, compliance with the AIA and GDPR is not merely a regulatory burden; it is a strategic imperative for building public trust, enabling cross-border scalability, and fostering sustainable innovation.

To move forward, Romania must invest in legal harmonisation, institutional reform, and practical support mechanisms for both public and private actors. Only through a coordinated, forward-looking approach can the country effectively integrate AI into its legal and economic systems — while upholding fundamental rights and participating meaningfully in the European digital transformation.

Romania, as a Member State of the European Union, must align itself with the use of AI technologies and, following the adoption of the AI Act, must

unquestionably adapt its entire legislative framework to reflect the aforementioned reference legal instruments. Clearly, at the institutional level, the real challenge lies in the effective mechanisms for responding to this complex set of requirements — responding proactively and comprehensively in areas where fundamental rights are affected by AI, in the context of the interplay between the AI Act and the GDPR, avoiding interpretative contradictions or overlapping regulations, and ensuring the availability of mechanisms both for contesting and for remedying decisions, as such situations will undoubtedly arise.

This “challenge” falls primarily on the Romanian legislator, who must be able to anticipate, at the national level, not only what has already emerged from the European framework, but also what is likely to arise from jurisprudence developed in parallel by the Court of Justice of the European Union (CJEU) and the European Court of Human Rights (ECHR). The legislator must gain a deep, evolving understanding of the new legal relationships that will continuously emerge from the already ongoing interaction between citizens and the legal entities developing AI technologies.

Bibliography

1. Al-Dulaimi, Ahmed Oudah Mohammed & Mohammed Abd-Al Wahab Mohammed. 2025 „Legal responsibility for errors caused by artificial intelligence (AI) in the public sector”. *International Journal of Law and Management*, <https://doi.org/10.1108/IJLMA-08-2024-0295>.
2. Bataineh, Abdallah Q., Alaa S. Mushtaha, Ibrahim A. Abu-AlSondos, Saeed Hameed Aldulaimi & Marwan Abdeldayem. 2024. "Ethical & Legal Concerns of Artificial Intelligence in the Healthcare Sector," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, Manama, Bahrain, pp. 491-495, doi: 10.1109/ICETISIS61505.2024.10459438.
3. Castro, Daniel and Michael McLaughlin, “Who Is Winning the AI Race: China, the EU, or the United States?” Center for Data Innovation, January 2021, <https://datainnovation.org/2021/01/who-is-winning-the-ai-race-china-the-eu-or-the-united-states-2021-update/>.
4. Chambers and Partners. (2024). Artificial Intelligence 2024 – Romania: Law & Practice Guide. Available at: <https://practiceguides.chambers.com/practice-guides/artificial-intelligence-2024/romania> [Accessed 21 Mar. 2025].
5. Enqvist, Lena 2024. "Rule-based versus AI-driven benefits allocation: GDPR and AIA legal implications and challenges for automation in public social security administration." *Information & Communications Technology Law* vol. 33, no. 2: 222-246, doi: 10.1080/13600834.2024.2349835.
6. Goktas, Polat. 2024. “Ethics, Transparency, and Explainability in Generative Ai Decision-Making Systems: A Comprehensive Bibliometric Study.” *Journal of Decision Systems*, October, 1–29. doi: 10.1080/12460125.2024.2410042.
7. Olimid, Anca Parmena, Catalina Maria Georgescu and Daniel Alin Olimid. 2024. "Legal Analysis of EU Artificial Intelligence Act (2024): Insights from

- Personal Data Governance and Health Policy." *Access to Justice in Eastern Europe* 7(4): 120-42 <<https://doi.org/10.33327/AJEE-18-7.4-a000103>>.
8. Quintais, João Pedro. 2025. "Generative AI, copyright and the AI Act." *Computer Law & Security Review*, vol. 56: 106107, <https://doi.org/10.1016/j.clsr.2025.106107>.
 9. Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
 10. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).
 11. Wallace, Nick and Daniel Castro (2018). *The Impact of the EU's New Data Protection Regulation on AI*. Information Technology and Innovation Foundation (ITIF). Available at: <https://itif.org/publications/2018/03/26/impact-eu-new-data-protection-regulation-ai> [Accessed 21 Mar. 2025].
 12. Wang, Xukang, Ying Cheng Wu, Xueliang Ji & Hongpeng Fu. 2024. "Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices." *Frontiers in Artificial Intelligence*, vol. 7: 1320277, <https://doi.org/10.3389/frai.2024.1320277>.
 13. Yuan, Yiming, Yongming Sun and Hangyu Chen. 2024. "Does Artificial Intelligence Affect Firms' Inner Wage Gap?" *Applied Economics* 57 (19): 2365–71. doi: 10.1080/00036846.2024.2324090.

The Use of Artificial Intelligence in Combating Tax Evasion: Challenges, Opportunities, and Ethical Implications from a Legal Perspective

Lecturer **Mihai ȘTEFĂNOAIA**¹

Abstract

Tax evasion is a persistent challenge for governments worldwide, leading to significant revenue losses and undermining public trust in fiscal systems. The integration of artificial intelligence (AI) into tax compliance and enforcement mechanisms presents a transformative opportunity to enhance detection and prevention capabilities. AI-driven tools, such as machine learning algorithms and predictive analytics, can identify fraudulent patterns, automate audits, and improve regulatory oversight². However, the adoption of AI in taxation also raises significant legal and ethical concerns, including data privacy, algorithmic bias, and due process rights³. From a legal standpoint, ensuring transparency and accountability in AI-based tax enforcement is crucial to maintaining fairness and preventing potential abuses. This paper explores the challenges, opportunities, and ethical dilemmas associated with AI-driven tax enforcement, analyzing regulatory frameworks and proposing legal safeguards for responsible AI implementation.

Keywords: artificial intelligence, tax evasion, legal implications, ethical challenges, regulatory frameworks.

JEL Classification: K22, K24, K34

DOI: <https://doi.org/10.62768/ADJURIS/2025/3/10>

Please cite this article as:

Ștefănoaia, Mihai, „The Use of Artificial Intelligence in Combating Tax Evasion: Challenges, Opportunities, and Ethical Implications from a Legal Perspective”, in Devetzis, Dimitrios, Dana Volosevici & Leonidas Sotiropoulos (eds.), *Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order*, ADJURIS – International Academic Publisher, Bucharest, Paris, Calgary, 2025, p. 178-189.

¹ Mihai Ștefănoaia - Faculty of Law and Administrative Sciences, „Ștefan cel Mare” University of Suceava, Romania, ORCID: 0000-0002-7163-4436, stefanoaiamihai@yahoo.com.

² Benjamin Alarie, *AI and the Future of Tax Avoidance* (December 4, 2023). Tax Notes Federal, December 4, 2023, p. 1809, Available at SSRN: <https://ssrn.com/abstract=4667814>.

³ Nuryani Nuryani, Achmad Benny Mutiara, I Made Wiryana, Detty Purnamasari, Souza Nurafrianto Windiartono Putra, (2024). „Artificial Intelligence Model for Detecting Tax Evasion Involving Complex Network Schemes”. *Aptisi Transactions on Technopreneurship (ATT)*, 6(3), 339–356. <https://doi.org/10.34306/att.v6i3.436>.

1. Introduction

The rapid technological evolution has fundamentally transformed how tax administrations collect, process, and analyze taxpayer data, leading to a significant transition from traditional tax audit methods — primarily based on manual declarations and spot checks — to integrated digital systems and advanced predictive analytics. The massive digitalization of economies, the rise of e-commerce, the globalization of financial flows, and the diversification of income sources have created unprecedented challenges for tax authorities. In this context, artificial intelligence (AI)-based tools have emerged as innovative solutions capable of managing vast volumes of data from heterogeneous sources and uncovering complex patterns of tax evasion that remain inaccessible through conventional means⁴.

AI is not merely a tool for automating repetitive tax processes but serves as a strategic partner in strengthening tax oversight capabilities. Through machine learning algorithms and predictive analytics, AI can identify tax evasion behaviors by correlating declared data with information obtained from external sources such as commercial registers, banking transactions, online platforms, and social networks⁵. For instance, by employing clustering techniques, AI systems can group taxpayers with similar behaviors and quickly detect deviations from typical profiles. This approach not only enhances the efficiency of tax audits but also enables proactive tax evasion prevention by signaling risks at an early stage.

Moreover, AI-driven systems contribute to increased voluntary compliance, as taxpayers become aware of the authorities' ability to detect irregularities swiftly, reducing the temptation to evade tax obligations⁶. The implementation of tax nudging systems — based on AI-driven personalized messaging — has proven to have positive effects on compliance by tailoring communication to each taxpayer's behavioral profile⁷.

However, the rapid expansion of AI use in the tax sphere raises multiple legal and ethical concerns that require appropriate and up-to-date regulation. From a legal perspective, the massive processing of taxpayer data raises questions about compliance with the principles of legality, proportionality, and the right to defense. For example, to what extent is the use of opaque and difficult-

⁴ OECD. (2025). *Tax Administration Digitalisation and Digital Transformation Initiatives*, <https://doi.org/10.1787/c076d776-en>.

⁵ European Commission (2022). *Artificial Intelligence Act: Proposal for a Regulation laying down harmonised rules on artificial intelligence*. COM/2021/206 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206>.

⁶ OECD (2021). *Tax Administration 3.0: The Digital Transformation of Tax Administration*. Paris: OECD Publishing, https://www.oecd.org/en/publications/tax-administration-3-0-the-digital-transformation-of-tax-administration_ca274cc5-en.html.

⁷ James Alm & Benno Torgler (2011). „Do Ethics Matter? Tax Compliance and Morality”. *Journal of Business Ethics*, 101, 635–651, <https://doi.org/10.1007/s10551-011-0761-9>.

to-explain algorithms compatible with the obligation to justify administrative tax decisions, as required by national and European legislation⁸? At the same time, AI may amplify the risks of indirect discrimination if algorithms are trained on historical datasets that reflect systemic biases or unequal treatment applied to certain categories of taxpayers.

From an ethical perspective, the balance between tax collection efficiency and the protection of fundamental rights — especially privacy and the presumption of innocence — becomes a critical issue. Intensive tax profiling, which effectively turns every taxpayer into a permanent suspect, contradicts the principles of a democratic rule-of-law state and risks undermining public trust in tax administrations⁹. Thus, AI integration into tax processes should not only be a technological modernization effort but also an opportunity to enhance transparency, accountability, and respect for taxpayer rights.

Therefore, it is essential that this technological revolution is accompanied by a corresponding adaptation of the legal framework, establishing clear limits for AI use in tax administration, algorithmic audit mechanisms, and effective safeguards for fundamental rights protection. Without such an integrated approach, the risk that AI becomes an abusive control tool — at the expense of tax fairness and social justice — remains high¹⁰.

The research questions are as follows:

1) What are the main legal challenges associated with the use of artificial intelligence in detecting and preventing tax evasion at national and international levels?

2) To what extent do artificial intelligence-based technologies improve the efficiency of tax authorities in combating tax evasion, and what are their implications for taxpayers' rights?

3) What ethical and data protection considerations should be taken into account in regulating the use of artificial intelligence for combating tax evasion?

4) How can the use of artificial intelligence in combating tax evasion be balanced with the principles of the rule of law and tax justice?

2. Opportunities Offered by AI in Combating Tax Evasion

The application of artificial intelligence in the tax domain creates multiple opportunities, including:

⁸ Christopher Barth Kuner, Daniel Cooper, 2017. *Data Protection Law and International Dispute Resolution*. Leiden/Boston: Brill - Nijhoff, 2017. 174 p. (Recueil des Cours: Collected Courses of the Hague Academy of International Law, Vol. 382), p. 78.

⁹ ECHR, case of S. and Marper v. the United Kingdom, 2008.

¹⁰ Jeffrey Owens, Ivan Lazarov and Nathalia Oliveira Costa, (2021), *Exploring the opportunities and challenges of new technologies for EU tax administration and policy*. European Parliament, [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695458/IPOL_STU\(2021\)695458_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695458/IPOL_STU(2021)695458_EN.pdf).

Automatic detection of fraud patterns. The use of artificial intelligence in combating tax evasion provides significant opportunities by automating analytical processes, increasing the accuracy of investigations, and reducing the time required to detect and prevent tax fraud. Beyond the automatic detection of fraud patterns, AI contributes to optimizing transaction monitoring, improving tax compliance, and enhancing the efficiency of control bodies.

Predictive analytics for tax evasion prevention. AI can be used to develop predictive models that anticipate evasion behaviors based on historical data and taxpayer characteristics. Machine learning algorithms can estimate the probability that a company or an individual will commit tax fraud and help authorities allocate control resources more efficiently¹¹. This approach enables proactive tax evasion prevention, thereby reducing financial losses to the state.

Real-Time transaction monitoring. AI-based technologies allow real-time monitoring of financial transactions and the detection of suspicious activities¹². By integrating data from multiple sources, such as tax declarations, bank transfers, and card payments, intelligent systems can automatically flag atypical or structured transactions designed to avoid taxation. For example, AI can detect the intentional fragmentation of payments to evade tax obligations or identify transactions between high-risk entities¹³.

Automation of tax audits and inspections. The use of AI in tax audit processes can significantly reduce the time needed to analyze documents and identify discrepancies. Natural language processing (NLP) systems can quickly examine large volumes of financial documents, extracting essential information for tax verifications¹⁴. Additionally, algorithms can prioritize cases with the highest risk of tax evasion, allowing tax inspectors to focus on the most relevant files¹⁵.

Enhancing tax compliance through virtual assistants. AI-powered virtual assistants can guide taxpayers in the compliance process, reducing errors and ambiguities in tax filings. These assistants can provide personalized recommendations, explanations of tax legislation, and early warnings in case of possible inconsistencies in tax declarations. Such intelligent support contributes to reducing unintentional evasion and improving tax transparency.

¹¹ OECD (2021). *Tax Administration 3.0: The Digital Transformation of Tax Administration*.

¹² Chirag Vinalbhai Shah, *Real-Time Transaction Monitoring: Combining AI, Big Data, and Biometric Authentication for Secure Payments*, June 2021, *Global Networks* 5(6): 38-47, DOI: 10.70179/GRDJEV09I100013.

¹³ IMF Annual Report, 2023, <https://www.imf.org/external/pubs/ft/ar/2023/english/>, accessed on 25.03.2025.

¹⁴ Friedrich Schneider, Andreas Buehn. "Shadow Economy: Estimation Methods, Problems, Results and Open questions" *Open Economics*, vol. 1, no. 1, 2018, pp. 1-29. <https://doi.org/10.1515/openec-2017-0001>

¹⁵ European Commission. (2022). Proposal for a Council Directive amending Directive 2006/112/EC as regards VAT rules for the digital age, Brussels, 8.12.2022 COM(2022) 701 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0701>.

Integration of blockchain with ai for increased transparency. The combination of AI and blockchain technology can significantly improve the traceability of financial transactions and prevent the manipulation of accounting data¹⁶. Blockchain provides an immutable ledger of transactions, while AI can analyze this data to detect anomalies and potential tax avoidance strategies. This synergy enables the creation of a more transparent and secure tax system.

Combating fraud in e-commerce and the digital economy. AI can be used to monitor online commercial activities and detect businesses that do not properly report their revenues¹⁷. By analyzing payment flows and online reviews, algorithms can identify merchants engaged in undeclared economic activities, thereby contributing to reducing tax evasion in the digital economy.

These opportunities highlight the potential of artificial intelligence to transform tax administration and significantly reduce losses caused by tax evasion. However, implementing such solutions requires appropriate regulation and close collaboration between tax authorities, financial institutions, and the private sector.

3. Legal and Technical Challenges in Applying Artificial Intelligence in the Tax Domain

The implementation of artificial intelligence in combating tax evasion and tax administration raises numerous challenges from both legal and technical perspectives. These difficulties stem from the need to balance the efficiency of automated processes with the protection of taxpayers' fundamental rights and ensuring the fairness of administrative decisions.

Protection of personal data and confidentiality. The processing of tax data using artificial intelligence must comply with the European legal framework on data protection, particularly Regulation (EU) 2016/679 (GDPR), which imposes strict restrictions on the collection, storage, and use of personal information¹⁸. The use of AI in analyzing tax data involves accessing massive databases, which can be correlated with external sources such as banks, social media, or commercial registries. This integration raises the risk of excessive surveillance and may lead to violations of the right to privacy, guaranteed by Article 8 of the Treaty on the Functioning of the European Union (TFEU). Furthermore, there is a risk that automated systems could retain data for longer periods than necessary or use it in ways incompatible with the original purpose of collection, which could result in legal sanctions for tax authorities.

¹⁶ OECD (2021). *Tax Administration 3.0: The Digital Transformation of Tax Administration*.

¹⁷ IMF Annual Report, 2023, <https://www.imf.org/external/pubs/ft/ar/2023/english/>, accessed on 25.03.2025.

¹⁸ European Data Protection Board. *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*, https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en.

Transparency and explainability of algorithms. A fundamental principle of administrative law is the justification of administrative acts, meaning that tax decisions must be clear and accessible to taxpayers¹⁹. When an AI algorithm is used to identify tax evasion risks or automatically generate tax assessments, taxpayers must have the ability to understand the logic behind the algorithm and contest any potential errors. However, many advanced machine learning models, especially those based on neural networks, function as “black boxes,” making it difficult to explain their outcomes in a comprehensible manner for end users²⁰. A lack of transparency may lead to legal challenges of tax decisions, arguing that the principle of legality and the right to a fair trial (Charter of Fundamental Rights of the EU, Article 47) have been violated. In this context, there is an urgent need to develop algorithmic audit mechanisms and establish standards for the explainability of models used in the tax domain.

Algorithmic discrimination and impact on tax equity. AI systems rely on historical data to learn patterns and make predictions; however, this data may contain errors, distortions, or systemic biases²¹. The uncontrolled application of algorithmic models in tax risk assessment could lead to discrimination against certain categories of taxpayers. For example, if historical data shows a higher rate of tax fraud in specific economic sectors or geographic regions, AI could automatically label SMEs in these areas as having a higher risk of evasion, leading to disproportionate tax audits. This situation contradicts the principle of equality before the law, enshrined in Article 20 of the Charter of Fundamental Rights of the EU. Additionally, AI algorithms could exacerbate existing inequalities in the tax system by favoring large taxpayers, who have the resources to legally optimize their tax obligations, while increasing the pressure on smaller taxpayers who do not have the same capacity for compliance²².

Legal liability for automated decisions. Another problematic aspect is determining legal liability when an AI algorithm makes an erroneous decision that affects a taxpayer. Currently, tax legislation does not provide a clear framework for assigning responsibility in such situations: who is responsible for an incorrect tax assessment generated by an automated system — the tax authority, the software developer, or the operator managing the algorithm? In the absence of specific regulations, taxpayers may face difficulties in challenging AI-based decisions, which could undermine their access to justice and protection of their

¹⁹ Christopher Barth Kuner, Daniel Cooper, *op. cit.*, 2017, p. 76.

²⁰ Sandra Wachter, Brent Mittelstadt, Luciano Floridi, 2017. „Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation”, *International Data Privacy Law*, Volume 7, Issue 2: 76–99, <https://doi.org/10.1093/idpl/ixp005>.

²¹ Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. „A Survey on Bias and Fairness in Machine Learning”. *ACM Computing Surveys* 54, 6, Article 115 (July 2022), 35 pages. <https://doi.org/10.1145/3457607>.

²² Solon Barocas, Moritz Hardt and Arvind Narayanan (2019). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, p. 50 et seq.

rights²³.

Cybersecurity and the risk of attacks on ai-based systems. The use of artificial intelligence in tax administration involves the integration of advanced data analytics systems; however, this exposes the tax infrastructure to significant cybersecurity risks. AI algorithms may be vulnerable to adversarial attacks, where malicious actors manipulate input data to induce errors in the system's predictions²⁴. For example, a taxpayer could attempt to alter the structure of their transactions to avoid detection by anti-fraud algorithms. Additionally, the use of AI requires storing large volumes of tax data, which increases the risk of cyberattacks and the leakage of confidential information.

Harmonization of tax legislation with technological advancements. Tax legislation is generally rigid and tailored to traditional tax collection mechanisms, which can pose an obstacle to the integration of emerging technologies. The rapid pace of technological progress makes it difficult to update the regulatory framework in a way that ensures both the efficiency of tax administration and the protection of taxpayers' rights. For example, many countries lack clear regulations regarding the use of AI in tax decision-making processes, which can create legal uncertainties and hinder the widespread adoption of these technologies.

In conclusion, the implementation of artificial intelligence in combating tax evasion offers considerable benefits but also raises multiple legal and technical challenges that require a balanced approach. Developing appropriate regulatory frameworks is essential to ensure transparency, equity, and the protection of taxpayers' rights without compromising the efficiency of tax administration.

4. Ethical and Legal Implications of Using Artificial Intelligence in the Tax Field

The use of artificial intelligence (AI) in tax monitoring and administration generates multiple ethical and legal implications that need to be addressed in order to ensure the protection of taxpayers' rights and the legality of the use of these technologies. From the issue of legal liability for erroneous decisions to the risk of excessive profiling, the use of AI must be accompanied by clear regulatory and oversight mechanisms to prevent abuses and ensure a balance between administrative efficiency and the protection of fundamental rights.

Legal liability for algorithmic errors. A major challenge in the use of AI in the tax field is determining liability when algorithms generate incorrect decisions or harm taxpayers. Errors can result from several factors, including

²³ OECD (2021). *Tax Administration 3.0: The Digital Transformation of Tax Administration*.

²⁴ Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik, Ananthram Swami (2017). *Practical Black-Box Attacks Against Machine Learning*. Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, UAE, 1(1), 45-57, <https://doi.org/10.48550/arXiv.1602.02697>.

faulty training data, imperfect algorithmic models, or misinterpretation of results. The key question is who should be held accountable: the tax authority using the technology, the software developer, or the technology oversight authority?

Currently, the legal framework in most EU member states does not provide clear regulations on liability in such cases²⁵. According to the general principle of administrative liability, tax authorities should assume responsibility for the use of AI, including for errors made. However, this principle does not cover situations where errors stem from the technical limitations of algorithms or from decisions made autonomously by advanced machine learning systems. Additionally, the lack of a clear algorithm audit framework complicates identifying the causes of erroneous decisions, which may hinder taxpayers' ability to challenge them.

One possible solution would be the adoption of a specific legal mechanism for algorithmic liability, similar to that used in the field of artificial intelligence applied to autonomous vehicles. This could include the obligation for tax authorities to demonstrate that AI decisions are correct and in compliance with current legislation, as well as the right of taxpayers to request human review of automated decisions²⁶.

Proportionality of ai use in relation to taxpayers' rights. A fundamental principle of European law is the principle of proportionality, which mandates that any administrative measure must be necessary, appropriate, and not excessively infringe upon the fundamental rights of the individuals concerned²⁷. In the tax context, the use of AI must be justified by a clear necessity and should not exceed what is strictly necessary to achieve the intended purpose.

The European Court of Human Rights has repeatedly emphasized the importance of balancing tax security with respect for privacy. For example, in the case of *S. and Marper v. the United Kingdom* (2008), the ECtHR ruled that the long-term retention of personal data without clear justification constitutes a violation of Article 8 of the European Convention on Human Rights, which protects privacy. This ruling can be interpreted as a relevant precedent for the use of AI in tax monitoring, suggesting that authorities cannot retain and analyze taxpayer data without clear and proportional justification.

In this regard, excessive use of AI in tax analysis could lead to unjustified intrusions into taxpayers' private lives. For example, the use of algorithms to analyze spending history, banking transactions, or even data from social media could violate the necessity and data minimization principles imposed by the GDPR. Therefore, it is essential that the implementation of AI in the tax field be accompanied by control and oversight mechanisms to ensure the proportionality

²⁵ Jeffrey Owens, Ivan Lazarov and Nathalia Oliveira Costa, *op. cit.* (2021), p. 70.

²⁶ Wachter, S., Mittelstadt, B., & Floridi, L., *op. cit.* (2017), p. 77.

²⁷ Oleksandr Kutovyi, Sergii Burma (2025). „Artificial Intelligence in the Case-Law of the European Court of Human Rights”. *Ehrlich's Journal*, (12), 34–44. <https://doi.org/10.32782/ehrlchsjournal-2025-12.05>.

of the measures applied.

Preventing “Excessive profiling”. AI-based tax profiling involves analyzing large volumes of data to identify behavior patterns and estimate the likelihood that a taxpayer is engaged in fraudulent activities. While this technology can be useful in detecting fraud, it raises serious concerns regarding the presumption of innocence and the risk of discriminatory treatment.

According to Article 48 of the Charter of Fundamental Rights of the EU, everyone is presumed innocent until proven guilty. Using AI to create detailed tax profiles, correlated with behavioral analysis and predictive models, could lead to treating taxpayers as suspects before an actual violation of the law is proven²⁸. For example, if an algorithm identifies a taxpayer as having a high risk of tax evasion based on factors such as the type of economic activity or geographic location, this could lead to more frequent tax inspections without the individual having committed fraud.

Another risk is posed by potential algorithmic errors that may affect specific groups of taxpayers. For instance, if historical data suggests that certain industries have a higher degree of tax evasion, algorithms may automatically label small businesses in those sectors as higher-risk, even without concrete evidence. This could lead to discriminatory application of tax measures and a violation of the principle of equality before the law.

To prevent these risks, legislation should impose clear limits on the use of AI in tax profiling. For example, algorithms should be periodically audited to identify potential biases, and taxpayers should have the right to contest decisions based on predictive models. Additionally, the use of AI for fully automated tax decisions, without human intervention, should be prohibited to ensure the respect of the right to a defense.

5. Conclusions

The use of artificial intelligence in tax monitoring and combating tax evasion offers significant advantages, but also involves major legal and ethical risks. The issue of liability for algorithmic errors, the need to comply with the proportionality principle, and the risk of excessive profiling are just a few of the challenges that need to be addressed to ensure fair and legally compliant implementation. A clear and transparent regulatory framework for the use of AI in the tax field is essential to prevent abuses and protect taxpayers' rights.

In the present paper, I have addressed the formulated research questions.

1. *What are the main legal challenges associated with the use of artificial intelligence in detecting and preventing tax evasion at national and international*

²⁸ Alessia Fidelangeli, Federico Galli (2021), „Artificial Intelligence and Tax Law: Perspectives and Challenges”, *Rivista Interdisciplinare Sul Diritto Delle Amministrazioni Pubbliche*, Fascicolo 4: 24-58. DOI: 10.13130/2723-9195/2021-4-27.

levels? is covered in Section III, Legal and technical challenges in applying artificial intelligence in the tax domain, where aspects such as data protection, algorithm transparency, algorithmic discrimination, legal liability, and cybersecurity are discussed.

2. *To what extent do artificial intelligence-based technologies improve the efficiency of tax authorities in combating tax evasion, and what are their implications for taxpayers' rights?* is examined in Section II. Opportunities offered by AI in combating tax evasion, where the benefits of AI, including automated fraud detection, predictive analytics, real-time monitoring, and the automation of tax inspections, are analyzed. Additionally, the implications for taxpayers' rights are discussed in Section III.

3. *What ethical and data protection considerations should be taken into account in regulating the use of artificial intelligence for combating tax evasion?* is analyzed in Section IV. Ethical and legal implications of using artificial intelligence in the tax field, where issues such as legal liability for algorithmic errors, the principle of proportionality, and the risk of excessive profiling are addressed.

4. *How can the use of artificial intelligence in combating tax evasion be balanced with the principles of the rule of law and tax justice?* is discussed in Sections III and IV, where the necessity of decision-making transparency, the prevention of algorithmic discrimination, and the protection of fundamental taxpayer rights are examined.

The implementation of artificial intelligence in combating tax evasion presents significant opportunities but also raises complex legal and ethical challenges. AI enhances the efficiency of tax authorities by automating fraud detection, improving predictive analytics, and enabling real-time transaction monitoring. However, its application must be carefully regulated to avoid infringements on fundamental rights.

One of the primary legal challenges concerns data protection and privacy, as AI systems require access to large tax-related datasets that may include sensitive personal information. Ensuring compliance with GDPR and other data protection laws is crucial to prevent excessive surveillance and unauthorized data use. Moreover, algorithmic transparency remains a major concern, as many AI systems operate as "black boxes," making it difficult for taxpayers to challenge tax assessments based on automated decisions. To align AI applications with the rule of law, tax authorities must implement clear accountability mechanisms and provide taxpayers with the right to human review of AI-driven tax decisions.

Ethically, profiling and algorithmic bias pose risks of discrimination and unequal tax enforcement. AI systems trained on historical tax data may inadvertently reinforce biases against certain economic sectors or demographic groups, leading to unfair tax audits and assessments. Therefore, regulatory frameworks must establish safeguards against discriminatory outcomes and mandate regular audits of AI models.

From a legal standpoint, establishing liability for AI-generated tax decisions is essential. Current tax regulations do not clearly define whether responsibility lies with tax authorities, software developers, or third-party AI providers. Without clear accountability, taxpayers may face difficulties contesting erroneous decisions, undermining their access to justice.

Furthermore, the integration of AI with blockchain technology offers a potential solution to improve financial transparency and tax compliance. Blockchain provides an immutable ledger of transactions, while AI can detect anomalies and flag potential tax avoidance schemes. However, this integration requires international legal harmonization to ensure cross-border enforcement and data-sharing regulations.

In conclusion, while AI has the potential to revolutionize tax administration by increasing efficiency and reducing fraud, its deployment must be balanced with legal safeguards to uphold transparency, fairness, and taxpayer rights. A robust regulatory framework is needed to ensure that AI-driven tax enforcement adheres to the principles of proportionality, non-discrimination, and legal accountability. Additionally, continuous collaboration between tax authorities, policymakers, and AI developers is necessary to refine legal standards and address emerging challenges in AI-powered tax compliance.

Bibliography

1. Alarie, Benjamin, 2023, *AI and the Future of Tax Avoidance (December 4, 2023)*. Tax Notes Federal, December 4, p. 1809, Available at SSRN: <https://ssrn.com/abstract=4667814>.
2. Alm, James & Benno Torgler (2011). „Do Ethics Matter? Tax Compliance and Morality”. *Journal of Business Ethics*, 101, 635-651, <https://doi.org/10.1007/s10551-011-0761-9>.
3. Barocas, Solon, Moritz Hardt and Arvind Narayanan (2019). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press.
4. European Commission (2022). *Artificial Intelligence Act: Proposal for a Regulation laying down harmonised rules on artificial intelligence*. COM/ 2021/206 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206>.
5. European Commission. (2022). *Proposal for a Council Directive amending Directive 2006/112/EC as regards VAT rules for the digital age*, Brussels, 8.12.2022 COM(2022) 701 final, <https://eur-lex.europa.eu/legal-content/EN/XT/PDF/?uri=CELEX:52022PC0701>.
6. European Data Protection Board. *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*, https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en.
7. Fidelangeli, Alessia & Federico Galli (2021), „Artificial Intelligence and Tax Law: Perspectives and Challenges”, *Rivista Interdisciplinare Sul Diritto Delle Amministrazioni Pubbliche*, Fascicolo 4: 24-58. DOI: 10.13130/2723-9195/20

- 21-4-27.
8. IMF Annual Report, 2023, <https://www.imf.org/external/pubs/ft/ar/2023/english/>, accessed on 25.03.2025.
9. Kuner, Christopher Barth & Daniel Cooper, 2017. *Data Protection Law and International Dispute Resolution*. Leiden/Boston: Brill - Nijhoff, 2017. 174 p. (Recueil des Cours: Collected Courses of the Hague Academy of International Law, Vol. 382).
10. Kutovyi, Oleksandr & Sergii Burma (2025). „Artificial Intelligence in the Case-Law of the European Court of Human Rights”. *Ehrlich's Journal*, (12), 34–44. https://doi.org/10.32782/ehrlc_hsjournal-2025-12.05.
11. Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. „A Survey on Bias and Fairness in Machine Learning”. *ACM Computing Surveys* 54, 6, Article 115 (July 2022), 35 pages. <https://doi.org/10.1145/3457607>.
12. Nuryani, Nuryani, Achmad Benny Mutiara, I Made Wiryana, Detty Purnamasari & Souza Nurafrianto Windiartono Putra, (2024). „Artificial Intelligence Model for Detecting Tax Evasion Involving Complex Network Schemes”. *Aptisi Transactions on Technopreneurship (ATT)*, 6(3), 339–356. <https://doi.org/10.34306/att.v6i3.436>.
13. OECD (2021). *Tax Administration 3.0: The Digital Transformation of Tax Administration*. Paris: OECD Publishing, https://www.oecd.org/en/publications/tax-administration-3-0-the-digital-transformation-of-tax-administration_ca274cc5-en.html.
14. OECD. (2025). *Tax Administration Digitalisation and Digital Transformation Initiatives*, <https://doi.org/10.1787/c076d776-en>.
15. Owens, Jeffrey, Ivan Lazarov and Nathalia Oliveira Costa, (2021), *Exploring the opportunities and challenges of new technologies for EU tax administration and policy*. European Parliament, [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695458/IPOL_STU\(2021\)695458_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695458/IPOL_STU(2021)695458_EN.pdf).
16. Papernot, Nicolas, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik & Ananthram Swami (2017). *Practical Black-Box Attacks Against Machine Learning*. Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, UAE, 1(1), 45-57, <https://doi.org/10.48550/arXiv.1602.02697>.
17. Schneider, Friedrich & Andreas Buehn. "Shadow Economy: Estimation Methods, Problems, Results and Open questions" *Open Economics*, vol. 1, no. 1, 2018, pp. 1-29. <https://doi.org/10.1515/openec-2017-0001>.
18. Shah, Chirag Vinalbhai, *Real-Time Transaction Monitoring: Combining AI, Big Data, and Biometric Authentication for Secure Payments*, June 2021, *Global Networks* 5(6): 38-47, DOI: 10.70179/GRDJEV09I100013.
19. Wachter, Sandra, Brent Mittelstadt & Luciano Floridi, 2017. „Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation”, *International Data Privacy Law*, Volume 7, Issue 2: 76–99, <https://doi.org/10.1093/idpl/ipy005>.