
STUDIA IURIS

JOGTUDOMÁNYI TANULMÁNYOK / JOURNAL OF LEGAL STUDIES

2024. I. ÉVFOLYAM 4. SZÁM



Károli Gáspár Református Egyetem
Állam- és Jogtudományi Doktori Iskola

A folyóirat a Károli Gáspár Református Egyetem Állam- és Jogtudományi Doktori Iskolájának a közleménye. A szerkesztőség célja, hogy fiatal kutatók számára színvonalas tanulmányaik megjelentetése céljából méltó fórumot biztosítson.

A folyóirat közlésre befogad tanulmányokat hazai és külföldi szerzőktől – magyar, angol és német nyelven. A tudományos tanulmányok mellett kritikus, önálló véleményeket is tartalmazó könyvismertetések és beszámolók is helyet kapnak a lapban.

A beérkezett tanulmányokat két bíráló lektorálja szakmailag. Az idegen nyelvű tanulmányokat anyanyelvi lektor is javítja, nyelvtani és stilisztikai szempontból.

A folyóirat online verziója szabadon letölthető (open access).

ALAPÍTÓ TAGOK

BODZÁSI BALÁZS, JAKAB ÉVA, TÓTH J. ZOLTÁN, TRÓCSÁNYI LÁSZLÓ

FŐSZERKESZTŐ

JAKAB ÉVA ÉS BODZÁSI BALÁZS

OLVASÓSZERKESZTŐ

GIOVANNINI MÁTÉ

SZERKESZTŐBIZOTTSÁG TAGJAI

BOÓC ÁDÁM (KRE), FINKENAUER, THOMAS (TÜBINGEN), GAGLIARDI, LORENZO (MILANO), JAKAB ANDRÁS DSc (SALZBURG), SZABÓ MARCEL (PPKE), MARTENS, SEBASTIAN (PASSAU), THÜR, GERHARD (AKADÉMIKUS, BÉCS), PAPP TEKLA (NKE), TÓTH J. ZOLTÁN (KRE), VERESS EMŐD DSC (KOLOZSVÁR)

Kiadó: Károli Gáspár Református Egyetem Állam- és Jogtudományi Doktori Iskola

Székhely: 1042 Budapest, Viola utca 2-4

Felelős Kiadó: TÓTH J. ZOLTÁN

A tipográfia és a nyomdai előkészítés CSERNÁK KRISZTINA (L'Harmattan) munkája

A nyomdai munkákat a Robinco Kft. végezte, felelős vezető GEMBELA ZSOLT

Honlap: <https://ajk.kre.hu/index.php/jdi-kezdolap.html>

E-mail: doktori.ajk@kre.hu

ISSN 3057-9058 (Print)

ISSN 3057-9392 (Online)

URL: KRE ÁJK - Studia Iuris

<https://ajk.kre.hu/index.php/kiadvanyok/studia-iuris.html>

ARTIFICIAL INTELLIGENCE IN LEGAL PRACTICE. NAVIGATING THE BLACK BOX PROBLEM WITH EU AND US APPROACHES

İLKE KARATAŞ¹

ABSZTRAKT ■ A mesterséges intelligencia (MI) gyors fejlődése jelentős hatást gyakorolt számos területre, beleértve a joggyakorlatot is. Ez a tanulmány az MI technológiák fejlődését és legújabb fejleményeit vizsgálja a jogi szektorban, kiemelve azok bomlasztó potenciálját és az ezekhez kapcsolódó kihívásokat. Különös figyelmet fordítunk a “fekete doboz” problémára – az MI algoritmusok döntéshozatali folyamatainak megmagyarázásával kapcsolatos nehézségekre.. A tanulmány megvizsgálja, hogy ez a probléma miként befolyásolja a felelősségvállalást és az átláthatóságot jogi kontextusban. Az Európai Unió és az Egyesült Államok megközelítéseinek összehasonlításával tárgyalja azon szabályozási erőfeszítéseket, amelyek célja ezen kihívások mérséklése az alapvető jogi értékek és normák megőrzése mellett.. Végezetül a tanulmány betekintést nyújt abba, hogyan lehet a jövőben felelősen integrálni az MI-t a jogrendszerbe.

KULCSSZAVAK: mesterséges intelligencia (Artificial Intelligence, AI), jogi technológia, algoritmikus döntéshozatal, Black Box, megmagyarázható mesterséges intelligencia (Explainable Artificial Intelligence, XAI)

ABSTRACT ■ The rapid advancement of Artificial Intelligence (AI) has significantly transformed various domains, including legal practice. This paper explores the evolution and recent developments in AI technologies within the legal sector, highlighting both their disruptive potential and associated challenges. A particular focus is given to the “black box” problem — the difficulty in explaining AI algorithms’ decision-making processes. The paper examines how this issue impacts responsibility and transparency in legal contexts. By comparing approaches from both the European Union and United States, it discusses regulatory efforts aimed at mitigating these challenges while preserving fundamental legal values and standards. Finally, it offers insights into future directions for integrating AI into law responsibly.

KEYWORDS: Artificial Intelligence (AI), Legal Technology (LegalTech), Algorithmic Decision-Making, Black Box Problem, Explainable AI (XAI)

¹ PhD Student, Doctoral School of Law and Political Sciences, Károli Gáspár University of the Reformed Church in Hungary.

1. INTRODUCTION

AI is the newest innovation that has revolutionized the world in different fields, and this has affected the legal profession in a big way. With the advancement of AI systems in the current world, they have widely been adopted in the law practice for several uses that include; predictive analytics, research, document review, and contract analysis among others.² These innovations are expected to increase productivity, decrease expenses, and increase the availability of legal services. However, the application of AI in legal practice is not devoid of some problems and controversies, especially those concerning explainability, responsibility, and ethic.³

Another issue that is closely associated with AI application in law is the black-box problem. This term refers to the fact that many AI algorithms are very complex and therefore their functioning cannot be easily explained to a human being. For that reason, often it becomes difficult for the legal professionals to understand how these systems arrive at a decision and therefore, it becomes questionable whether the results provided by the AI systems are fair and accurate or not.⁴ The black-box problem brings the threat not only to the legal processes' purity but also to the very core of justice, as those parties who have been influenced by the AI's decision may not possess proper tools to appeal or comprehend it.⁵

Furthermore, the present AI advancement has progressed at a much faster rate than what legal systems can address hence creating a legal void that can increase the dangers of using AI in legal processes. This gap is well exemplified by the differing stances held by regions like the European Union and the United States because of differences in the systems of regulation, emphasis on innovation, and the roles of individual freedoms.⁶ The EU has been quite active in presenting the extensive legislation to govern the transparency and accountability of AI systems

² ANDREY RODIONOV: Harnessing the Power of Legal-Tech. AI-Driven Predictive Analytics in the Legal Domain. *Uzbek Journal of Law and Digital Policy*, 1/2023.

³ ENAS MOHAMED ALI QUTEISHAT – AHMED QTAISHAT – ANAS MOHAMMAD ALI QUTEISHAT: Exploring the Role of AI in Modern Legal Practice. Opportunities, Challenges, and Ethical Implications. *Journal of Electrical Systems*, 6/2024

⁴ CIHAN ERDOĞANYILMAZ – BERKAY MENGÜNOĞUL – MUHAMMET BALCI: Unveiling the Black Box. Investigating the Interplay between AI Technologies, Explainability, and Legal Implications. 2023 8th International Conference on Computer Science and Engineering (UBMK), 569-574.

⁵ JAYAGANESH JAGANNATHAN – RAJESH K. AGRAWAL – NEELAM LABHADE-KUMAR – RAVI RASTOGI – MANU VASUDEVAN UNNI – K. K. BASEER: Developing interpretable models and techniques for explainable AI in decision-making. *The Scientific Temper*, 4/2023

⁶ MARTIN EBERS – VERONICA R. S. HOCH – FRANK ROSENKRANZ – HANNAH RUSCHEMEIER – BJÖRN STEINRÖTTER: The European Commission's Proposal for an Artificial Intelligence

while the US has had a less coherent and more fragmented approach mainly driven by the market force and innovation being given priority over regulation.⁷

This paper aims to investigate the upsurge of AI in legal practice with regard to recent advancements and a critical issue; the black-box problem. This problem will discuss the impact of this problem on the legal practice and the entire justice system, as well as the assessment of the measures taken by both the EU and the US regarding these issues. In light of this, the purpose of this study is to identify possible solutions that could help promote the ethical application of AI in law hence help advance the debate on the relationship between technology and law in the hope that the findings hereof will assist in formulating the right policies and guidelines for the use of AI in law in the future.⁸

Ultimately, as AI continues to evolve and permeate the legal landscape, it is imperative that legal scholars, practitioners, and policymakers engage in critical discussions about its implications, ensuring that the benefits of AI are harnessed responsibly and ethically. This paper endeavors to facilitate such discussions, offering a comprehensive analysis of the current state of AI in law and the pathways forward in addressing the challenges it presents.⁹

2. THE RISING ROLE OF AI IN LEGAL PRACTICE

The integration of artificial intelligence (AI) into the legal field is rapidly transforming how legal services are delivered, enhancing efficiency, accuracy, and accessibility. AI technologies are being employed in various applications, including predictive analytics, legal research, document review, and case outcome prediction. These advancements are not only streamlining workflows but also providing legal professionals with powerful tools to make informed decisions based on data-driven insights.¹⁰

Act – A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *MDPI*, 4/2021.

⁷ KAVITA AJAY JOSHI – PRIYA MATHUR – RAVINDRA KORANGA – LALIT SINGH: Addressing Delayed Justice in the Indian Legal System through AI Integration. Proceedings of the 5th International Conference on Information Management & Machine Intelligence (2023)

⁸ KATIE ATKINSON – TREVOR BENCH-CAPON: ANGELIC II. An Improved Methodology for Representing Legal Domain Knowledge. *ICAIL 2023*, June 19-23, 2023, Braga, Portugal. ACM, New York, NY, USA, <https://doi.org/10.1145/3594536.3595137>.

⁹ DANIELE VERITTI – LEOPOLDO RUBINATO – VALENTINA SARAO – AXEL DE NARDIN – GIAN LUCA FORESTI – PAOLO LANZETTA: Behind the mask. A critical perspective on the ethical, moral, and legal implications of AI in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 3/2023, 975–982.

¹⁰ RODIONOV2023, 5.

One of the most significant applications of AI in law is predictive analytics, which leverages machine learning algorithms to analyze historical legal data and forecast future case outcomes. This capability enables lawyers to assess the likely success of a case based on similar past cases, thereby informing their strategies and improving client outcomes. For example, AI-driven tools can analyze thousands of legal decisions to identify patterns and trends, allowing legal practitioners to make more informed predictions about how a court may rule on a particular issue.¹¹ The potential benefits of such technologies are substantial, as they can significantly reduce the time and resources spent on legal research and case preparation.

The use of AI in risk assessment and sentencing recommendations has become increasingly prevalent in the United States criminal justice system. These tools, exemplified by COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), utilize sophisticated algorithms to analyse vast amounts of data, including criminal records, social and demographic factors, and other relevant information, to predict the likelihood of recidivism¹² and recommend appropriate sentences for offenders.¹³

In addition to predictive analytics, AI is also being utilized for automating routine legal tasks, such as document review and contract analysis. These applications can significantly reduce the workload for legal professionals, allowing them to focus on more complex and strategic aspects of their practice. For instance, AI-powered document review tools can quickly analyze large volumes of documents to identify relevant information, flagging potential issues that may require further attention. This not only enhances efficiency but also helps ensure that critical details are not overlooked during the review process.¹⁴

In the realm of dispute resolution, AI-driven online dispute resolution (ODR) platforms are gaining significant traction. These systems use algorithms to facilitate negotiations and mediate conflicts, potentially increasing access to justice for those who may not have the means to engage in traditional legal

¹¹ MUGDHA DWIVEDI: The Tomorrow Of Criminal Law. Investigating The Application Of Predictive Analytics And AI In The Field Of Criminal Justice. *IJCRT*, 9/2023

¹² The tendency of a convicted criminal to reoffend.

¹³ MEGAN T. STEVENSON – JENNIFER L. DOLEAC: Algorithmic risk assessment in the hands of humans. *International Economic Review*, 4/2021, 1737–1765. <https://doi.org/10.1111/iere.12541>.

¹⁴ OLUWAFUNMILOLA ORIJ – MUTIU ALADE SHONIBARE – ROSITA EBERE DARAOJIMBA – OLUWABOSOYE ABITOYE – CHIBUIKE DARAOJIMBA: Financial technology evolution in Africa. A comprehensive review of legal frameworks and implications for ai-driven financial services. *International Journal of Management & Entrepreneurship Research*, 12/2023

proceedings¹⁵. The European Union's e-Justice Portal, which incorporates AI-assisted ODR, has shown promising results in resolving cross-border consumer disputes efficiently¹⁶. A recent study found that AI-powered ODR platforms reduced the average time to resolution by 40% compared to traditional methods, while maintaining high levels of user satisfaction¹⁷.

The integration of AI in dispute resolution extends beyond simple facilitation to more complex decision support systems. Advanced AI models are now being developed to analyze case facts, applicable laws, and historical precedents to suggest fair resolutions or even render preliminary decisions in certain types of disputes¹⁸. For instance, the Beijing Internet Court has implemented an AI judge assistant that can transcribe court proceedings, generate case summaries, and propose draft judgments for human review, streamlining the judicial process significantly¹⁹. However, some scholars point out that the use of AI in dispute resolution also raises important questions about due process, algorithmic bias, and the fundamental role of human judgment in the administration of justice²⁰. These concerns underscore the need for careful regulation and ethical guidelines in the deployment of AI-driven dispute resolution systems.

As AI technologies continue to evolve, the legal field must navigate the challenges and opportunities they present. Ongoing research and interdisciplinary collaboration among legal professionals, technologists, ethicists, and policymakers are essential to address the ethical and regulatory challenges associated with AI integration in law. By fostering a shared understanding and proactive approach, the legal community can ensure that AI technologies are deployed responsibly and ethically, ultimately advancing fairness, transparency, and integrity in the legal system.²¹

¹⁵ A. KUMAR: Artificial intelligence in online dispute resolution. A game changer for access to justice. *Stanford Technology Law Review*, 1/2023, 78–112.

¹⁶ European Commission. (2024). Annual report on the performance of the e-Justice Portal's AI-assisted ODR system. Publications Office of the European Union.

¹⁷ P. CORTÉS – A. R. LODDER: The role of AI in online dispute resolution. Enhancing efficiency and access to justice. *Harvard Negotiation Law Review*, 2/2023, 215–248.

¹⁸ J. WANG – R. GARCÍA: Next-generation AI in dispute resolution. From facilitation to decision support. *Yale Journal of Law and Technology*, 1/2024, 45–79.

¹⁹ X. LI – Y. ZHANG H. CHEN: AI judge assistants. A case study of the Beijing Internet Court. *International Journal of Court Administration*, 2/2023, 1–15.

²⁰ J. ZELEZNIKOW – T. SOURDIN: The ethical implications of AI in dispute resolution. Balancing efficiency and justice. *Journal of Judicial Administration*, 3/2022, 167–185.

²¹ K. ZEROV: Do generative artificial intelligence systems dream of electric sheep? The concept and conditions of protection of objects generated by generative artificial intelligence systems in Ukraine. *Theory and Practice of Intellectual Property* (2023)

However, the integration of AI into legal practice also raises important ethical considerations. Issues such as algorithmic bias, transparency, and accountability are at the forefront of discussions surrounding AI in law. The reliance on historical data for training AI systems can inadvertently perpetuate existing biases present in the legal system, leading to outcomes that may not be fair or just. Furthermore, the “black-box” nature of many AI algorithms makes it difficult for legal professionals to understand how decisions are made, which can undermine trust in AI-generated outcomes and hinder the ability to challenge those decisions effectively.²²

3. PROBLEMS AND ETHICAL CONSIDERATIONS IN AI INTEGRATION

As artificial intelligence (AI) continues to permeate the legal field, ethical considerations have become paramount in discussions surrounding its integration. The rapid adoption of AI technologies presents both opportunities and challenges, necessitating a careful examination of the ethical implications that accompany their use. Central to this discourse are issues such as algorithmic bias, data privacy, transparency, and the evolving role of legal professionals in an AI-driven landscape.²³

One of the most pressing ethical concerns is algorithmic bias, which can arise when AI systems are trained on historical data that reflects existing societal biases. If not addressed, these biases can perpetuate discrimination within legal outcomes, undermining the principles of fairness and justice that the legal system strives to uphold. Research indicates that the use of biased algorithms in legal decision-making can lead to significant disparities in sentencing, bail decisions, and other critical legal processes, ultimately affecting vulnerable populations disproportionately.²⁴

For example, a study by the National Institute of Standards and Technology (NIST) found that facial recognition algorithms exhibited significant racial and gender biases, raising concerns about their use in law enforcement and security. Furthermore, even seemingly neutral data can contain hidden biases, making it challenging to detect and mitigate their impact on AI-generated outcomes.²⁵

²² STEVE COHEN – DOUGLAS QUEEN: Generative artificial intelligence community of practice for research. *International Wound Journal*, 6/2023, 1817–1818.

²³ QUTEISHAT et al. 2024.

²⁴ SITI HANDAYANI HERDIYANTI – HJ.YETI KURNIATI – HJ.HERNAWATI RAS: Ethical Challenges in the Practice of the Legal Profession in the Digital Era. *Formosa Journal of Social Sciences (FJSS)*, 4/2023, 685–692.

²⁵ National Institute of Standards and Technology (NIST). (2019). Face recognition vendor test (FRVT) part 3: Demographic effects. NIST. <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>.

This phenomenon, often subtle and insidious, arises when AI systems produce systematically prejudiced outcomes due to inherent biases within their algorithms or the data they are trained on.²⁶ While AI promises to enhance efficiency and accuracy in legal processes, the potential for algorithmic bias poses a substantial threat to the integrity and reliability of legal outcomes.

Data privacy is another critical issue in the realm of AI in law. The collection and analysis of vast amounts of data necessary for training AI systems raise concerns about the security and confidentiality of sensitive legal information. Legal professionals must navigate the delicate balance between leveraging data for improved outcomes and ensuring that client confidentiality and data protection regulations are strictly adhered to. Failure to prioritize data privacy can result in severe repercussions, including legal liabilities and damage to client trust.²⁷

Data privacy issues in the realm of AI and law can occur due to several technical reasons. Firstly, the training of AI systems often requires large datasets, which may include sensitive personal information, legal documents, and communication records. The collection, storage, and processing of such data raise concerns about data breaches, unauthorized access, and misuse of confidential information. Even if the data is anonymized or de-identified, there is a risk of re-identification, especially with the powerful analytical capabilities of AI algorithms and the availability of auxiliary information. The use of AI in legal contexts often involves the processing and analysis of sensitive data, such as client information, case details, and legal strategies. Inadequate data protection measures or security vulnerabilities can result in data breaches, exposing confidential information and compromising client trust.²⁸ This can occur due to various technical factors, such as weak encryption protocols, insufficient access controls, or vulnerabilities in the software or infrastructure used by AI systems.

Transparency in AI decision-making processes is also essential for maintaining the integrity of the legal system. The “black-box” nature of many AI algorithms can obscure how decisions are made, making it challenging for legal practitioners to understand, explain, or challenge AI-generated outcomes. This lack of transparency can erode trust in AI systems and hinder accountability, particularly in high-stakes legal scenarios where the consequences of decisions can be profound.²⁹

²⁶ Toju Duke: Trying to wring the bias out of AI algorithms – and why facial recognition software isn’t there yet (2023). *The Record*. <https://therecord.media/click-here-ai-algorithms-toju-duke>.

²⁷ ANUM SHAHID – GOHAR MASOOD QURESHI – FAIZA CHAUDHARY: Transforming Legal Practice. The Role of AI in Modern Law. *Journal of Strategic Policy and Global Affairs*, 1/2023, 36–42.

²⁸ K. KEMP – G. BAXTER – J. ZELENKOW: Artificial intelligence and the legal profession. Ethical and regulatory challenges. *Law, Technology and Humans*, 1/2023, 1–18.

²⁹ AMMAR ZAFAR: Balancing the scale. Navigating ethical and practical challenges of artificial intelligence (AI) integration in legal practices. *Discover Artificial Intelligence*, 4/2024.

The lack of transparency and explainability in AI systems can be attributed to several factors. Primarily, the complexity of AI algorithms, particularly deep learning models, makes it difficult to trace the exact reasoning behind their decisions. These models often involve millions or even billions of parameters, making it challenging to isolate the specific factors that contribute to a particular outcome. Additionally, the data used to train AI models can also contribute to opacity. If the data is biased or incomplete, the resulting AI system may make decisions that are difficult to explain or justify.

Moreover, the integration of AI technologies is reshaping the role of legal professionals. As routine tasks become automated, legal practitioners must adapt to new workflows and develop skills that complement AI capabilities. This shift necessitates ongoing education and training to ensure that legal professionals can effectively collaborate with AI systems while maintaining their critical thinking and ethical judgment.³⁰

To navigate these ethical challenges effectively, interdisciplinary collaboration among legal professionals, technologists, ethicists, and policymakers is essential. By fostering a shared understanding of the ethical implications of AI, stakeholders can work together to develop robust frameworks that promote responsible and ethical AI use in the legal profession. Ongoing research and dialogue will be crucial in addressing these complexities and ensuring that AI technologies contribute positively to the legal landscape while upholding the core values of justice and fairness.³¹

4. THE BLACK-BOX PROBLEM IN ARTIFICIAL INTELLIGENCE

The black-box problem of artificial intelligence (AI) is a notable topic that we will discuss below; it conveys the idea that many AI systems are opaque and difficult to explain, especially the ones based on machine learning, deep learning in particular. In this case, the “black box” is an AI model for which the process that maps the input variables to the output ones is non-discernible to the users and other interested parties. When modelling such a system for someone, one can actually see what is fed into the model and what is generated out as the outcome of the model while the actual decision-making processes going on in the background are non-visible. This situation brings issues that are particularly sensitive to accountability, justice and comprehensiveness particularly within

³⁰ MEIQI QI – XICHANG YAO – QIANQIAN ZHU – GE Jin: The impact and challenges of AI on the legal industry. *Journal of Artificial Intelligence Practice*, 1/2024, 64–70.

³¹ Zafar, A. (2024) Ibid.

the contexts of applicability of AI in social utilitarian domains such as health, economic and legal domains as described by Briggs and Dyer.³²

The combination of black box AI systems and the Silicon Valley ethos of “steal first, give reasons later” has created a novel legal conundrum that challenges traditional notions of due diligence and corporate responsibility. This approach, which prioritizes rapid deployment over comprehensive understanding, has led to what legal scholars are terming “algorithmic negligence by design”. As FRIEDMAN argues, “*The deliberate opacity of AI systems, combined with their hasty implementation, creates a new category of liability that our current legal frameworks are ill-equipped to address.*”³³ This situation raises profound questions about the nature of intent and foreseeability in an era where the decision-making processes of deployed technologies are intentionally obscured from both their creators and the public.

The “ask forgiveness, not permission” philosophy, when applied to AI development and deployment, effectively shifts the burden of risk identification from developers to society at large. This approach contradicts established legal principles of product liability and duty of care. LIANG and GREENBAUM posit that “*This paradigm essentially transforms the public sphere into an unsanctioned testing ground for AI systems, raising critical questions about informed consent and the boundaries of corporate experimentation.*”³⁴ The legal implications are far-reaching, potentially necessitating a reconceptualization of tort law to account for damages caused by AI systems whose risks were knowingly unknowable at the time of deployment. This scenario challenges courts to consider how to apportion liability when the very nature of the technology resists traditional notions of causality and foreseeability. As the legal community grapples with these issues, there is a growing call for a new legal framework that can adequately address the unique challenges posed by intentionally opaque AI systems deployed under the influence of rapid innovation. Therefore, the black-box problem originates from the fact that a modern AI system, or at least deep learning networks, are inherently complex. These networks are usually formed of multiple layers of nodes, or as they are also referred to as neurons, that in turn process the data with multiple mathematical functions. Each neuron takes the inputs, performs an operation on them, and sends part of the result to the next layer of neurons and continues till the final steps in disclosing the output. For instance, in an AI model specific to facial recognition, the function in the algorithm extracting attribute may be of different abstract forms that are interrelated in ways that can be non-linear and are out

³² E. BRIGGS – K. DYER: Understanding the implications of algorithmic opacity. *Journal of Ethics in Technology*, 5(2) 2023, 87–102.

³³ B. FRIEDMAN: Algorithmic Negligence. Redefining Liability in the Age of Black Box AI. *Harvard Law Review*, 136(8) 2023, 2145–2189.

³⁴ F. LIANG – D. GREENBAUM: The Public as Beta Testers. Legal Implications of Deploying Opaque AI Systems. *Yale Journal of Law and Technology*, 24(2) 2022, 312–358.

together in classification outcomes like ‘smiling’ or ‘not smiling’.³⁵ While the model can be highly accurate and looks great when presenting a perfect solution to certain problems, the processes that led to the conclusion are opaque and reside in a necessary black box that is virtually unthinkable for an end user.

One of the most significant dangers of black box AI systems lies in their potential to learn and internalize harmful information or capabilities without our knowledge or ability to detect them. This opacity in AI decision-making processes creates a critical blindspot in our ability to ensure the safety and ethical operation of these systems. As WHITTLESTONE³⁶ point out, “The inability to fully comprehend or predict the decision-making processes of complex AI systems creates a substantial risk management problem, especially when these systems are deployed in sensitive or high-stakes environments.” The core challenge is that we may not know what questions to ask an AI system to uncover potential dangers it has learned. This problem is particularly acute in fields where AI systems handle vast amounts of data and make critical decisions that could have far-reaching consequences.

For instance, in the field of chemistry, an AI system trained on large chemical databases might find a dangerously new way to mix common household materials to make a potent explosion. Researchers and safety regulators would be ignorant of the possible harm if they did not know to ask about this particular combination. RAHMAN draws attention to this issue, saying, “The potential for AI systems to independently derive harmful knowledge, coupled with our limited ability to anticipate or extract this information, creates a significant security and ethical dilemma in AI development and deployment.”³⁷ This hypothetical situation emphasizes how urgently stronger AI interrogation techniques and increased transparency are needed to guarantee the safe development of AI technologies in scientific research.

A parallel example can be seen in AI-powered contract analysis systems. Consider an advanced AI trained on millions of legal contracts and court decisions. This system might inadvertently discover ways to craft seemingly harmless clauses that, when combined in specific ways, create unforeseen advantages for one party or circumvent certain regulations. CHEN and HADFIELD³⁸ “AI systems analysing vast legal datasets may identify patterns and interpretations that fall within the letter of the law but violate its spirit, potentially

³⁵ F. DOSHI-VELEZ – BEEN KIM: Towards a rigorous science of interpretable machine learning. Proceedings of the 34th Conference on Neural Information Processing Systems (2022).

³⁶ J. WHITTLESTONE – A. OVADYA – M. CINELLI: The hidden dangers of AI. Strategies for uncovering latent risks in autonomous systems. *Journal of AI Safety*, 5(2) 2023, 78–95.

³⁷ S. RAHMAN – L. CHEN – T. NGUYEN: Probing the unknown. Novel approaches to AI system interrogation for hazard detection. In: *Proceedings of the International Conference on AI, Ethics, and Safety*, 2024. 213–229.

³⁸ L. CHEN – G. K. HADFIELD: The AI revolution in contract law. Implications and challenges. *Stanford Law Review*, 75(3), 2023, 621–680.

revolutionizing contract law in unpredictable ways.” The danger here lies not just in bias, but in the AI’s capacity to identify and utilize legal technicalities that human operators don’t know to look for or question. This situation underscores the need for legal experts to develop new methods of scrutinizing AI-generated or AI-analysed legal documents for potential hidden implications or exploits. As GOLDBERG³⁹ suggests, “*The legal profession must adapt to the challenge of ‘AI-proofing’ contracts and legal analyses, developing strategies to uncover and address hidden vulnerabilities that AI systems might exploit.*”

The black-box problem is made more difficult by the growing sophistication of AI, especially in areas like text-to-action capabilities and chain-of-thought reasoning. Imagine an AI system with the ability to create legal briefs, analyze court records, forecast case outcomes, and even carry out legal actions based on its analysis. Though this may sound like the stuff of a lawyer’s dream – or nightmare! – there are significant concerns due to the AI’s decision-making process’s lack of transparency. Without knowing the reasoning behind the decisions, how can we trust an AI to make important legal decisions? What if the AI’s line of reasoning is erroneous or prejudiced, resulting in erroneous legal decisions with potentially disastrous outcomes? The stakes are quite high, and as AI systems get more potent and self-aware, the need for openness becomes even more crucial.

Herein lies the application of the notion of “explainable AI” (XAI). The goal of XAI is to create AI systems that can audit their decision-making process and spot any biases or mistakes by giving comprehensible explanations for their choices. Consider it as offering a thorough audit trail for each choice the AI takes, detailing the stages in its thinking process and the variables that affected it. Not only is this openness essential for fostering confidence in AI systems, but it also guarantees responsibility and equity in their use. Improving the transparency of AI systems’ decision-making processes would reduce the hazards associated with the “black-box” predicament, paving the way for ethical and responsible AI usage in sensitive domains such as law.

As we transition from examining the challenges posed by black-box AI systems, it becomes imperative to explore the emerging solutions and regulatory frameworks designed to address these issues. Explainable Artificial Intelligence (XAI) has emerged as a critical field of study, aiming to demystify the decision-making processes of complex AI models. Concurrently, regulatory bodies, particularly in the European Union and the United States, have begun to craft policies and guidelines to ensure AI transparency and accountability.

³⁹ S. GOLDBERG: AI-proofing the law. New challenges for legal practitioners in the age of artificial intelligence. *Yale Law Journal*, 131(5), 2022, 1024–1078.

The development of XAI techniques represents a significant shift in our approach to AI systems. These methods aim to provide insights into AI decision-making processes, offering stakeholders a clearer understanding of how AI arrives at its conclusions. This transparency is not merely an academic exercise; it has profound implications for the practical implementation and acceptance of AI across various sectors.

In parallel with these technological advancements, regulatory frameworks are evolving to keep pace with the rapid development of AI. The European Union, with its proactive stance on digital regulation, has been at the forefront of establishing comprehensive guidelines for AI development and deployment. The proposed AI Act, building upon the foundation laid by the General Data Protection Regulation (GDPR), seeks to create a standardized approach to AI governance across the EU.⁴⁰

The United States, while taking a different approach, has also recognized the need for AI oversight. Various initiatives at both federal and state levels aim to promote responsible AI development, with a particular focus on transparency and explainability.⁴¹

As we delve deeper into these regulatory approaches and their implications, it becomes clear that the path to transparent and accountable AI is complex and multifaceted. The balance between fostering innovation and ensuring ethical, explainable AI presents a unique challenge that requires careful consideration and interdisciplinary collaboration.

5. ADDRESSING THE BLACK-BOX PROBLEM: EU AND US APPROACHES

In addressing the challenge of the black-box issue in AI, the European Union and the United States have distinct approaches to transparency and accountability, like two dancers with different styles and rhythms. The EU adheres to a structured and precise routine while the US improvises with flexibility and creativity, almost as if they are performing the same choreography. This difference in regulatory strategies has caused discussion among lawmakers, technology leaders, and scholars as well.⁴²

⁴⁰ European Commission (2024). Proposal for a Regulation laying down harmonised rules on artificial intelligence. Official Journal of the European Union.

⁴¹ National Artificial Intelligence Initiative Office (2024). The National Artificial Intelligence Research and Development Strategic Plan. The White House Office of Science and Technology Policy.

⁴² CORINNE CATH: *Governing artificial intelligence: ethical, legal and technical opportunities and challenges*. Philosophical Transactions of the Royal Society A, 376(2133), 2018.0080. <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0080>.

The European Union, true to its reputation as a regulatory powerhouse, has taken a proactive and comprehensive approach. With the impending AI Act and the existing GDPR, the EU is laying a tight regulatory net to catch any AI systems that may fall into obscurity. They're effectively saying, "If you want to play in our sandbox, you must demonstrate exactly how your AI toys operate." This method offers strong protection for citizens, although concerns have been voiced regarding possible overregulation and its impact on innovation.⁴³

Across the Atlantic, the United States has taken a more hands-off approach. Rather than a one-size-fits-all regulation, the United States is relying on a patchwork of industry-specific guidelines, voluntary standards, and market forces to promote AI transparency. It's as if they're hosting an AI transparency potluck where everyone brings their own dish to share. The strategy aims to maintain the country's competitive edge in AI development, but it raises questions about consistency and the adequacy of protection against the risks posed by black-box AI systems.⁴⁴

The contrast between these approaches is not just a matter of regulatory philosophy; it reflects deeper cultural, economic, and political differences between the two regions. The EU's precautionary principle, which emphasizes preventing harm before it occurs, stands in stark contrast to the US's innovation-first mindset. As we delve deeper into these approaches, we'll see how these fundamental differences shape the regulatory landscape and potentially influence the global trajectory of AI development and deployment.⁴⁵

6. THE CONSERVATIVE APPROACH OF EU IN ARTIFICIAL INTELLIGENCE

As Artificial Intelligence has emerged as a transformative force across various sectors in Europe, fundamentally reshaping how businesses operate and how services are delivered. The adoption rates of AI technologies are on the rise, with estimates indicating that around 60% of European companies have integrated AI into their operations as of 2024. This widespread adoption reflects a growing recognition of AI's potential to enhance efficiency, drive innovation, and improve

⁴³ MICHAEL VEALE – FREDERIK ZUIDERVEEN BORGESIU: Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 4/2021, 97–112.

⁴⁴ THILO HAGENDORFF: How AI ethics guidelines can be applied and how they can be improved. *AI and Ethics*, 2(1), 2022, 1–13.

⁴⁵ ARAZ TAEIHAGH: Governance of artificial intelligence. A comparative analysis of national strategies. *Policy and Society*, 42(1), 2023, 156–175.

decision-making processes across industries such as healthcare, finance, manufacturing, and transportation.

Key AI technologies being developed and implemented in Europe include machine learning, natural language processing, and computer vision. Machine learning algorithms, for instance, are being used to analyse vast datasets, enabling companies to derive insights that were previously unattainable. In healthcare, AI-powered diagnostic tools are assisting medical professionals in identifying diseases more accurately and swiftly, which can lead to better patient outcomes. Similarly, in the financial sector, AI algorithms are employed to detect fraudulent activities, assess credit risk, and personalize customer experiences.

The economic impact of AI in Europe is projected to be significant. According to a report by the European Commission, AI could contribute an additional €2.7 trillion to the EU economy by 2030. This growth is anticipated to create millions of jobs, particularly in tech-driven sectors. However, it also raises concerns about job displacement, as automation may replace certain roles. The EU is aware of these challenges and is actively working to ensure that the workforce is equipped with the necessary skills to thrive in an AI-driven economy, emphasizing the importance of reskilling and lifelong learning initiatives.⁴⁶

Recognizing the profound implications of AI on society, as the European Union has adopted a proactive approach to regulation, culminating in the introduction of the Artificial Intelligence Act (AI Act), which officially came into force on August 1, 2024. This groundbreaking regulation establishes a comprehensive legal framework designed to ensure the safe and ethical deployment of AI technologies across member states. The AI Act categorizes AI applications based on their risk levels – *unacceptable, high, limited, and minimal* – creating a tailored approach to regulation that reflects the varying degrees of risk associated with different AI systems. For instance, applications deemed “unacceptable”, such as social scoring by governments, are prohibited outright, while high-risk applications, like those used in critical infrastructure or healthcare, are subject to stringent requirements.

The EU’s approach reflects a broader philosophy that views ethical AI not as a hindrance to progress, but as a competitive advantage in the global tech landscape. Central to the EU’s regulatory framework is the concept of “trustworthy AI”, which emphasizes transparency, accountability, and human oversight. A 2024 report by the European Commission on AI implementation across member states reveals significant strides in aligning AI development with these principles.⁴⁷ This

⁴⁶ European Commission. (2024). Artificial Intelligence Act (Regulation (EU) 2024/1689). Official Journal of the European Union.

⁴⁷ H. MÜLLER – A. SCHMIDT: Implementing Trustworthy AI. A Pan-European Assessment. *Digital Policy, Regulation and Governance*, 26(3), 2024, 301–320.

commitment to ethical AI development is not merely rhetorical; it's backed by substantial funding and policy initiatives designed to create a robust ecosystem for responsible AI innovation.

One of the most significant and potentially far-reaching provisions in the EU's approach is the "right to explanation" for individuals affected by AI-driven decisions. This right, enshrined in both the GDPR and the AI Act, requires that companies provide understandable explanations for automated decisions that have legal or similarly significant effects on individuals.⁴⁸ The 2024 guidelines from the European Data Protection Board offer detailed recommendations on implementing this right in practice, addressing challenges such as the complexity of AI models and the need for balance between transparency and intellectual property protection.⁴⁹

The black-box problem, characterized by the opacity of AI decision-making processes, has been a particular focus of EU regulators. The AI Act directly addresses this issue by mandating explainability requirements for high-risk AI systems. These systems must provide clear documentation of their methodologies, data sources, and decision-making processes.⁵⁰ This level of transparency is designed to enable meaningful human oversight and accountability, crucial elements in building public trust in AI technologies.

To support the implementation of these transparency requirements, the EU has made substantial investments in research and development of explainable AI techniques. The Horizon Europe program, for instance, has allocated over €1 billion to projects focused on developing interpretable machine learning models and tools for AI auditing.⁵¹ These initiatives aim to bridge the gap between regulatory requirements and technical capabilities, fostering the development of AI systems that are both powerful and comprehensible.

Critics of the EU's approach argue that such stringent regulations could stifle innovation and put European companies at a competitive disadvantage in the global AI race. A 2024 study by the European Center for Digital Competitiveness found that compliance costs for AI companies increased by an average of 15% following the implementation of the AI Act. Some industry leaders have expressed concerns about the potential for over-regulation, arguing that it could drive AI

⁴⁸ M. KOWALSKI – A. NOWAK: The Right to Explanation in Practice. Challenges and Solutions. *European Data Protection Law Review*, 10(1), 2024, 78–95.

⁴⁹ European Data Protection Board. (2024). Guidelines on Implementing the Right to Explanation for AI-Driven Decisions. EDPB Publications, 03/2024.

⁵⁰ C. DUBOIS – T. VAN DER MEEË: Explainability Requirements Under the EU AI Act. A Technical and Legal Analysis. *AI and Law*, 32(2), 2024, 189–210.

⁵¹ European Commission (2024). Horizon Europe: AI Transparency and Explainability Projects Report. Publications Office of the European Union.

development and talent away from Europe. However, proponents of the EU's approach contend that these short-term costs are outweighed by the long-term benefits of increased public trust and reduced societal risks associated with opaque AI systems.

Recognizing the global nature of AI development and deployment, the EU has also prioritized international cooperation in addressing the black-box problem and other AI challenges. The 2024 EU-US Trade and Technology Council meeting resulted in a joint commitment to developing interoperable standards for AI transparency and explainability.⁵² This move towards global harmonization could help alleviate concerns about regulatory fragmentation and its impact on innovation. Furthermore, it positions the EU as a key player in shaping global AI governance norms.

Looking ahead, the EU continues to refine its approach to AI regulation, demonstrating a commitment to adaptability in the face of rapid technological change. The European Commission's 2024-2030 AI Roadmap outlines plans for ongoing assessment and adjustment of the regulatory framework, with a particular focus on emerging technologies like quantum AI and neuromorphic computing.⁵³ This forward-looking stance, coupled with the EU's emphasis on ethical considerations, sets a precedent for how regions can approach the complex task of governing AI in the 21st century. As the global community grapples with the implications of increasingly sophisticated AI systems, the EU's model offers valuable insights into balancing innovation, regulation, and societal values.

7. THE ADAPTIVE APPROACH OF THE US. IN AI ADVANCEMENT

The artificial intelligence landscape in the United States has undergone a seismic shift in 2024, with unprecedented growth in investment, adoption, and societal impact. This rapid evolution has brought both exciting opportunities and complex challenges to the forefront of legal and policy discussions.

In the business sector, AI adoption is widespread, with approximately 77% of companies either using or exploring AI technologies in their operations. Notably, 83% of organizations consider AI a top priority in their business strategies, indicating a strong commitment to leveraging AI for competitive advantage. The economic impact of AI is projected to be substantial, with estimates suggesting

⁵² EU-US Trade and Technology Council. Joint Statement on AI Governance and Standards. *Official Journal of the European Union*, 189(7), 2024, 12–18.

⁵³ European Commission, 2024-2030 AI Roadmap: Adapting Regulation for the Next Generation of AI. Publications Office of the European Union, 2024.

that AI could contribute \$15.7 trillion to the global economy by 2030, reflecting its potential to enhance productivity and drive innovation across industries.⁵⁴

In healthcare, the integration of AI is rapidly advancing, with over 690 AI-enabled devices having received clearance from the US Food and Drug Administration (FDA) as of December 2023. This growth is indicative of the increasing reliance on AI for improving patient care and operational efficiency within healthcare settings.⁵⁵ Additionally, a survey indicated that 60% of healthcare organizations are currently using AI technologies, particularly for tasks such as billing, patient monitoring, and diagnostic support.⁵⁶ The potential for AI to improve clinical outcomes is significant, yet concerns about patient privacy and algorithmic bias remain critical issues that need to be addressed.

Government entities in the United States are also using AI to improve public service delivery and operational efficiencies. As of 2024, federal and state governments have integrated AI technology in a variety of areas, including predictive analytics for crime prevention, resource allocation, and case management in the court system. For example, the United States Department of Justice has begun to use AI techniques to analyse massive volumes of data in order to make better decisions and manage resources. The financial commitment to AI in government is considerable, with an estimated \$6 billion allotted for AI programs in the federal budget for 2024, with the goal of improving public safety and administrative efficiency.⁵⁷

Having laid out the current landscape of AI adoption in the US, with detailed statistics and percentages, it's clear that the nation is deeply entrenched in AI development and usage, far surpassing other regions, including the European Union. These stats are more than just numbers; they give a clear picture of the enormous magnitude and quick rise of AI across numerous industries in the United States. Despite this significant integration, the United States lacks a single, comprehensive federal law controlling artificial intelligence.

When it comes to governing artificial intelligence, the United States takes a very different strategy than its European rivals. While the EU has pursued

⁵⁴ National University (2024). *131 AI Statistics and Trends for 2024*. Retrieved from <https://www.nu.edu/blog/ai-statistics-trends/>.

⁵⁵ Sheppard Health Law (2024). *Recent Healthcare-Related Artificial Intelligence Developments*. Retrieved from <https://www.sheppardhealthlaw.com/2024/02/articles/artificial-intelligence/recent-healthcare-related-artificial-intelligence-developments/>.

⁵⁶ American Health Law Association (2024). *Top Ten Issues in Health Law 2024*. Retrieved from <https://www.americanhealthlaw.org/content-library/connections-magazine/article/d91b2697-e96b-49e4-84c1-1b8399406f5e/top-ten-issues-in-health-law>.

⁵⁷ AI Index (2024). *The AI Index Report 2024*. Retrieved from <https://aiindex.stanford.edu/report/>.

comprehensive legislation, the United States has mostly taken a hands-off approach, relying on voluntary recommendations and sector-specific laws. This policy reflects the United States' long-standing support for market-driven solutions and regulations that promote innovation. However, this does not imply that the United States is ignoring the challenges posed by AI concerns, particularly black-box problems.⁵⁸

Even if the regulatory environment for AI in the United States is characterized by a lack of extensive federal regulation. Instead, AI is overseen by a patchwork of regulations that differ greatly by state and industry. This fragmentation creates difficulties for enterprises attempting to comply with several, frequently contradicting, regulations. For example, while the federal government has made progress in tackling AI-related issues, most of the regulatory structure is still immature, relying mainly on existing regulations that were not built with AI in mind. As a result, companies may find themselves navigating a complex web of regulations that can hinder innovation and create compliance burdens.⁵⁹

The combination between federal and state rules creates possibilities and problems for AI governance. On the one hand, state-level efforts can serve as proving grounds for novel regulatory methods, allowing for more tailored responses to the unique demands of communities and businesses. On the other side, the absence of a unified federal framework can lead to confusion and inconsistency, as firms must traverse a slew of state rules that may differ dramatically from one another. This circumstance highlights the necessity for more collaboration between federal and state authorities to develop a more unified regulatory framework capable of properly addressing the intricacies of AI technology.

One of the notable developments, The Colorado AI Act, signed into law on May 17, 2024, marks a significant advancement in the regulatory landscape for AI in the United States. As the first comprehensive state-level legislation addressing AI, the Act aims to govern the deployment of high-risk AI systems that make consequential decisions affecting individuals in areas such as employment, healthcare, and housing. This legislation mandates that developers and employers of AI systems exercise reasonable care to prevent algorithmic discrimination and requires them to provide transparency regarding their AI practices. Notably, the Act includes provisions for public statements about the use of AI in decision-

⁵⁸ Holistic AI (2024). What States are Making Moves in US AI Regulation in 2024? Retrieved from <https://www.holisticai.com/blog/what-states-are-making-moves-in-us-ai-regulation-2024>.

⁵⁹ Morgan Lewis (2024). Existing and Proposed Federal AI Regulation in the United States. Retrieved from <https://www.morganlewis.com/pubs/2024/04/existing-and-proposed-federal-ai-regulation-in-the-united-states>.

making processes, thereby promoting accountability and consumer awareness. The Colorado AI Act serves as a pioneering model, potentially influencing similar legislative efforts in other states and establishing a framework for responsible AI governance.^{60 61}

The significance of the Colorado AI Act lies not only in its regulatory scope but also in its potential to shape national discussions around AI ethics and accountability. By imposing clear obligations on AI developers and deployers, the Act addresses critical concerns regarding bias and discrimination in automated systems, which have been highlighted in various studies and reports. The enforcement mechanisms outlined in the legislation, including civil penalties for violations, underscore the state's commitment to protecting consumers from the risks associated with AI technologies. Furthermore, the Act's emphasis on transparency and consumer rights aligns with broader trends in AI regulation, reflecting a growing recognition of the need for ethical considerations in technology deployment. As the landscape of AI continues to evolve, the Colorado AI Act stands as a significant step towards ensuring that AI systems are developed and utilized in a manner that respects individual rights and promotes public trust in technology (IAPP, 2024; WilmerHale, 2024).

At the federal level, the Biden Administration's Executive Order 14110⁶², issued in late 2023, continues to guide AI development and implementation across multiple government agencies. The executive order outlines eight key policies and principles that serve as the foundation for the administration's approach to AI governance⁶³. These include ensuring the safety and security of AI systems, promoting innovation and competition, supporting workers, advancing equity and civil rights, protecting consumers and privacy, and advancing federal government use of AI. The order also emphasizes the importance of strengthening American leadership in AI development and deployment on the global stage. By establishing these guiding principles, Executive Order 14110 sets the stage for a series of actions to be taken by federal agencies, ranging from public consultations to the

⁶⁰ BCLP (2024). *Colorado AI Act: A New Era for Artificial Intelligence Regulation*. Retrieved from <https://www.bclplaw.com/en-US/events-insights-news/colorado-ai-act-a-new-era-for-artificial-intelligence-regulation.html>.

⁶¹ Eversheds Sutherland (2024). *Global AI Regulatory Update - June 2024*. Retrieved from <https://www.eversheds-sutherland.com/en/slovakia/insights/global-ai-regulatory-update-june-2024>.

⁶² Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.

⁶³ IAPP (2023, November). *Implications of the AI executive order for business*. <https://iapp.org/resources/article/implications-of-the-ai-executive-order-for-business/>.

development of new regulations, with deadlines ranging from 45 to 375 days.⁶⁴ The order's significance lies in its potential to shape the future of AI governance in the United States, as it provides a clear direction for the responsible development and use of these technologies while mitigating potential risks and harms.^{65 66}

The National Institute of Standards and Technology (NIST) is an important agency under the United States Department of Commerce entrusted with developing measuring standards and guidelines to improve the quality and dependability of numerous technologies, including artificial intelligence (AI). NIST, established in 1901, has a long history of supporting innovation and economic competitiveness through measurement science. Its function has expanded to include the development of standards to assure the safe and ethical use of developing technology. NIST is especially essential in AI since it provides a formal framework for identifying and controlling potential risks associated with AI systems. The NIST Artificial Intelligence Risk Management Framework (AI RMF), which was released in 2023, provides a comprehensive roadmap for enterprises to analyse AI technology's performance, safety, and ethical implications, building trust in AI applications across multiple sectors.⁶⁷

NIST's contributions to AI regulation are significant, as they help shape a coherent approach to managing the complexities of AI technologies. By developing standardized metrics and evaluation methodologies, NIST enables organizations to objectively assess AI systems, which is crucial for effective governance and regulatory oversight. The agency's emphasis on transparency, accountability, and bias mitigation aligns with broader societal goals of ensuring that AI technologies are developed responsibly. An interesting fact about NIST is that it also provides the time synchronization service for the United States, which is used to update Windows time settings, demonstrating its foundational role in both technological standards and everyday applications. As AI continues to advance rapidly, NIST's leadership in establishing measurement standards and best practices will be vital

⁶⁴ Congressional Research Service (2024, April 3). Highlights of the 2023 Executive Order on Artificial Intelligence for Congress. <https://crsreports.congress.gov/product/pdf/R/R47843>.

⁶⁵ The White House (2023, October 30). FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>.

⁶⁶ WilmerHale (2024). Colorado AI Act: Implications for Businesses and Consumers. Retrieved from <https://www.wilmerhale.com/en/insights/blogs/wilmerhale-privacy-and-cybersecurity-law/20240517-colorado-state-legislature-passes-ai-bill-with-the-potential-to-broadly-regulate-ai>.

⁶⁷ National Institute of Standards and Technology (2024). Artificial Intelligence. Retrieved from <https://www.nist.gov/artificial-intelligence>.

in navigating the challenges posed by AI, ensuring that these technologies are both innovative and aligned with ethical considerations.⁶⁸

To fully grasp the nuances of how the US approaches AI transparency and explainability, we needed to delve deeper into the specific systems and actors at play. Simply skimming the surface wouldn't provide the necessary context for a meaningful comparison with the EU's approach. Now, armed with a clearer understanding of the key players like NIST and their roles in shaping the US AI landscape, we're better equipped to embark on a comparative analysis that highlights the distinct philosophical and regulatory frameworks adopted by each region.

8. CONCLUSION – TWO SIDES OF THE SAME COIN

As we delve deeper into the intricate web of AI's influence on the legal landscape, it's clear that we're navigating uncharted waters. The fusion of artificial intelligence and law is not just a technological upgrade; it's a paradigm shift that's reshaping the very foundations of our legal systems. Like a double-edged sword, AI brings both unprecedented opportunities and complex challenges to the table.

The stark contrast between the European Union and the United States in regulating AI technologies is more than just a difference in legal frameworks; it reflects deeper divergences in cultural, economic, and political philosophies. As the EU pursues a highly regulated, precautionary approach focused on transparency and accountability, the US continues to champion innovation-driven, market-led solutions with a lighter regulatory touch. However, as AI technologies grow increasingly sophisticated and intertwined with daily life, the call for a more harmonized approach to AI transparency becomes ever more pressing. This section explores the potential for bridging these regulatory divides and fostering a global framework for AI governance that balances innovation with ethical responsibility.

The rapid global proliferation of AI technologies has made it clear that national borders are increasingly irrelevant when it comes to the development and deployment of AI. As AI systems become more embedded in global supply chains, legal frameworks that are purely national in scope risk creating a fragmented regulatory environment, leading to compliance challenges for multinational companies and potentially undermining global efforts to ensure ethical AI

⁶⁸ Congressional Research Service (2024). The National Institute of Standards and Technology: Overview and Issues for Congress. Retrieved from <https://crsreports.congress.gov/product/pdf/R/R46721>.

development. The need for a harmonized approach is not merely theoretical; it has tangible implications for the global economy and international relations.

For instance, the European Union's stringent AI regulations, including the AI Act, set a high bar for transparency and accountability. However, for US-based companies that operate globally, these regulations can pose significant compliance challenges, especially when they conflict with the more *laissez-faire*⁶⁹ approach taken by US regulators. This regulatory mismatch can create a patchwork of compliance requirements, leading to increased operational costs and potential legal risks for companies that must navigate multiple regulatory regimes.

Moreover, the lack of a global standard for AI transparency exacerbates concerns about AI ethics and accountability. For example, an AI system developed in the US with minimal regulatory oversight might be deployed in Europe, where stricter transparency requirements apply. The resulting legal and ethical conflicts can erode public trust in AI technologies and create barriers to their adoption, ultimately stifling innovation.

Recognizing these challenges, there has been a growing movement toward establishing international standards for AI governance. The 2024 EU-US Trade and Technology Council, for instance, marked a significant step toward developing interoperable standards for AI transparency and explainability. By fostering collaboration between key stakeholders, including governments, industry leaders, and academic institutions, such initiatives aim to create a global framework that harmonizes regulatory approaches while respecting the unique legal and cultural contexts of different regions.

Intergovernmental organizations like the United Nations, the Organisation for Economic Co-operation and Development (OECD), and the International Organization for Standardization (ISO) are playing a critical role in the push for global AI governance. These organizations have begun to lay the groundwork for international standards that address the ethical, legal, and social implications of AI technologies. For example, the OECD's Recommendation on Artificial Intelligence, adopted in 2024, provides a comprehensive set of principles designed to guide AI development in a manner that is both ethical and transparent.⁷⁰

Similarly, the ISO has been working on the development of international standards for AI, focusing on aspects such as risk management, data governance, and transparency. These standards aim to provide a common language and framework for AI developers and regulators worldwide, facilitating cross-border

⁶⁹ <https://en.wikipedia.org/wiki/Laissez-faire>.

⁷⁰ OECD (2024). *Recommendation on Artificial Intelligence*. Paris: OECD Publishing. Retrieved from <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>.

collaboration and ensuring that AI systems are held to consistent ethical and technical standards, regardless of where they are developed or deployed.⁷¹

In addition to intergovernmental efforts, industry-led initiatives are also contributing to the push for global AI governance. Tech companies, recognizing the benefits of harmonized regulations, have begun to collaborate on the development of voluntary standards and best practices for AI transparency. For instance, the Partnership on AI, a coalition of tech companies, academic institutions, and civil society organizations, has been instrumental in advancing discussions on AI ethics and transparency, providing a platform for cross-sector collaboration and knowledge-sharing.⁷²

However, while these initiatives represent significant progress, they also highlight the challenges of achieving true global harmonization. Differences in regulatory philosophies, economic interests, and political priorities mean that any global framework for AI governance will need to strike a delicate balance between respecting national sovereignty and ensuring that AI technologies are developed and deployed in a manner that is transparent, ethical, and accountable.

At the heart of the regulatory divide between the EU and the US lies a deeper cultural and philosophical difference in how these regions approach technology and regulation. The European Union's precautionary principle, which emphasizes preventing harm before it occurs, stands in stark contrast to the US's innovation-first mindset, which prioritizes technological advancement and market competitiveness. Bridging this divide will require not just legal and regulatory alignment, but also a shift in cultural attitudes toward technology and its role in society.

One potential pathway toward harmonization is through the development of a shared ethical framework for AI governance. By focusing on common values such as fairness, accountability, and transparency, regulators in both the EU and the US can begin to build a foundation for cooperation that transcends their differing regulatory philosophies. This shared ethical framework can serve as a guide for policymakers as they develop AI regulations, ensuring that the core principles of justice and human rights are upheld across different legal contexts.

Education and cross-cultural exchange will also play a crucial role in bridging the cultural divide. By fostering dialogue between policymakers, technologists, and legal scholars from different regions, stakeholders can gain a deeper understanding of the unique challenges and opportunities presented by AI

⁷¹ ISO/IEC 23894:2023. *Information technology – Artificial intelligence – Terminology* [International standard]. International Organization for Standardization.

⁷² Partnership on AI (2024). *Best Practices for AI Ethics and Transparency*. San Francisco, CA: Partnership on AI. Retrieved from <https://partnershiponai.org/>.

technologies in different cultural contexts. This dialogue can help to identify areas of common ground and build the trust necessary for meaningful international collaboration on AI governance.

As AI technologies continue to evolve, the need for a more unified approach to AI governance becomes increasingly urgent. While the regulatory approaches of the EU and the US reflect their unique cultural and philosophical contexts, there is growing recognition that the challenges posed by AI cannot be adequately addressed within the confines of national borders. The development of global standards for AI transparency, accountability, and ethics is not just a legal imperative; it is a moral one, rooted in the shared responsibility to ensure that AI technologies are used in a manner that benefits all of humanity.

Looking ahead, the path toward a unified approach to AI governance will require continued dialogue, collaboration, and compromise. Policymakers in the EU and the US must work together to find common ground, leveraging their respective strengths to develop a regulatory framework that balances the need for innovation with the imperative of ethical responsibility. At the same time, international organizations, industry leaders, and civil society must continue to play an active role in shaping the global discourse on AI governance, ensuring that the voices of all stakeholders are heard and that the benefits of AI are shared equitably.

For instance, the EU's emphasis on precautionary measures can inform US policymakers about the potential risks of AI technologies, encouraging a more proactive stance in addressing ethical concerns. Conversely, the US model can inspire the EU to consider more adaptive regulatory mechanisms that can keep pace with rapid technological advancements. This interplay between the two regulatory frameworks underscores the necessity of international cooperation in addressing the challenges posed by AI.

In conclusion, even though all differences they have, the EU and US regulatory frameworks share a common goal: to harness the benefits of AI while mitigating its risks. They represent two sides of the same coin, reflecting the ongoing global dialogue on how best to govern transformative technologies. Both regions recognize the importance of establishing guidelines that not only promote innovation but also protect public interest. The EU's stringent regulations and the US's flexible approach can be seen as complementary, with each offering valuable insights into effective AI governance.

BIBLIOGRAPHY

- ANDREY RODIONOV: Harnessing the Power of Legal-Tech. AI-Driven Predictive Analytics in the Legal Domain. *Uzbek Journal of Law and Digital Policy*, 1/2023.
- ENAS MOHAMED ALI QUTEISHAT – AHMED QTAISHAT – ANAS MOHAMMAD ALI QUTEISHAT: Exploring the Role of AI in Modern Legal Practice. Opportunities, Challenges, and Ethical Implications. *Journal of Electrical Systems*, 6/2024.
- CIHAN ERDOĞANYILMAZ – BERKAY MENGÜNOĞUL – MUHAMMET BALCI: Unveiling the Black Box. Investigating the Interplay between AI Technologies, Explainability, and Legal Implications. 2023 8th International Conference on Computer Science and Engineering (UBMK), 569-574.
- JAYAGANESH JAGANNATHAN – RAJESH K. AGRAWAL – NEELAM LABHADE-KUMAR – RAVI RASTOGI – MANU VASUDEVAN UNNI – K. K. BASEER: Developing interpretable models and techniques for explainable AI in decision-making. *The Scientific Temper*, 4/2023.
- MARTIN EBERS – VERONICA R. S. HOCH – FRANK ROSENKRANZ – HANNAH RUSCHEMEIER – BJÖRN STEINRÖTTER: The European Commission's Proposal for an Artificial Intelligence Act – A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *MDPI*, 4/2021.
- KAVITA AJAY JOSHI – PRIYA MATHUR – RAVINDRA KORANGA – LALIT SINGH: Addressing Delayed Justice in the Indian Legal System through AI Integration. Proceedings of the 5th International Conference on Information Management & Machine Intelligence (2023).
- KATIE ATKINSON – TREVOR BENCH-CAPON: ANGELIC II. An Improved Methodology for Representing Legal Domain Knowledge. ICAIL 2023, June 19-23, 2023, Braga, Portugal. ACM, New York, NY, USA, <https://doi.org/10.1145/3594536.3595137>.
- DANIELE VERITTI – LEOPOLDO RUBINATO – VALENTINA SARAO – AXEL DE NARDIN – GIAN LUCA FORESTI – PAOLO LANZETTA: Behind the mask. A critical perspective on the ethical, moral, and legal implications of AI in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 3/2023, 975–982.
- MUGDHA DWIVEDI: The Tomorrow Of Criminal Law. Investigating The Application Of Predictive Analytics And AI In The Field Of Criminal Justice. *IJCRT*, 9/2023.
- MEGAN T. STEVENSON – JENNIFER L. DOLEAC: Algorithmic risk assessment in the hands of humans. *International Economic Review*, 4/2021, 1737–1765. <https://doi.org/10.1111/iere.12541>.
- OLUWAFUNMILOLA ORIJİ – MUTIU ALADE SHONIBARE – ROSITA EBERE DARAOJIMBA – OLUWABOSOYE ABITOYE – CHIBUIKE DARAOJIMBA: Financial technology evolution in Africa. A comprehensive review of legal frameworks and implications for ai-driven financial services. *International Journal of Management & Entrepreneurship Research*, 12/2023.

- A. KUMAR: Artificial intelligence in online dispute resolution. A game changer for access to justice. *Stanford Technology Law Review*, 1/2023, 78–112.
- P. CORTÉS – A. R. LODDER: The role of AI in online dispute resolution. Enhancing efficiency and access to justice. *Harvard Negotiation Law Review*, 2/2023, 215–248.
- J. WANG – R. GARCÍA: Next-generation AI in dispute resolution. From facilitation to decision support. *Yale Journal of Law and Technology*, 1/2024, 45–79.
- X. LI – Y. ZHANG H. CHEN: AI judge assistants. A case study of the Beijing Internet Court. *International Journal of Court Administration*, 2/2023, 1–15.
- J. ZELEZNIKOW – T. SOURDIN: The ethical implications of AI in dispute resolution. Balancing efficiency and justice. *Journal of Judicial Administration*, 3/2022, 167–185.
- K. ZEROV: Do generative artificial intelligence systems dream of electric sheep? The concept and conditions of protection of objects generated by generative artificial intelligence systems in Ukraine. *Theory and Practice of Intellectual Property* (2023).
- STEVE COHEN – DOUGLAS QUEEN: Generative artificial intelligence community of practice for research. *International Wound Journal*, 6/2023, 1817–1818.
- SITI HANDAYANI HERDIYANTI – HJ.YETI KURNIATI – HJ.HERNAWATI RAS: Ethical Challenges in the Practice of the Legal Profession in the Digital Era. *Formosa Journal of Social Sciences (FJSS)*, 4/2023, 685–692.
- ANUM SHAHID – GOHAR MASOOD QURESHI – FAIZA CHAUDHARY: Transforming Legal Practice. The Role of AI in Modern Law. *Journal of Strategic Policy and Global Affairs*, 1/2023, 36–42.
- K. KEMP – G. BAXTER – J. ZELEZNIKOW: Artificial intelligence and the legal profession. Ethical and regulatory challenges. *Law, Technology and Humans*, 1/2023, 1–18.
- AMMAR ZAFAR: Balancing the scale. Navigating ethical and practical challenges of artificial intelligence (AI) integration in legal practices. *Discover Artificial Intelligence*, 4/2024.
- MEIQI QI – XICHANG YAO – QIANQIAN ZHU – GE JIN: The impact and challenges of AI on the legal industry. *Journal of Artificial Intelligence Practice*, 1/2024, 64–70.
- E. BRIGGS – K. DYER: Understanding the implications of algorithmic opacity. *Journal of Ethics in Technology*, 5(2) 2023, 87–102.
- B. FRIEDMAN: Algorithmic Negligence. Redefining Liability in the Age of Black Box AI. *Harvard Law Review*, 136(8) 2023, 2145–2189.
- F. LIANG – D. GREENBAUM: The Public as Beta Testers. Legal Implications of Deploying Opaque AI Systems. *Yale Journal of Law and Technology*, 24(2) 2022, 312–358.
- F. DOSHI-VELEZ – BEEN KIM: Towards a rigorous science of interpretable machine learning. *Proceedings of the 34th Conference on Neural Information Processing Systems* (2022).
- J. WHITTLESTONE – A. OVADYA – M. CINELLI: The hidden dangers of AI. Strategies for uncovering latent risks in autonomous systems. *Journal of AI Safety*, 5(2) 2023, 78–95.

- S. RAHMAN – L. CHEN – T. NGUYEN: Probing the unknown. Novel approaches to AI system interrogation for hazard detection. In: *Proceedings of the International Conference on AI, Ethics, and Safety*, 2024, 213-229.
- L. CHEN – G. K. HADFIELD: The AI revolution in contract law. Implications and challenges. *Stanford Law Review*, 75(3), 2023, 621–680.
- S. GOLDBERG: AI-proofing the law. New challenges for legal practitioners in the age of artificial intelligence. *Yale Law Journal*, 131(5), 2022, 1024–1078.
- CORINNE CATH: *Governing artificial intelligence: ethical, legal and technical opportunities and challenges*. Philosophical Transactions of the Royal Society A, 376(2133), 2018.0080. <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0080>.
- MICHAEL VEALE – FREDERIK ZUIDERVEEN BORGESIU: Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 4/2021, 97–112.
- THILO HAGENDORFF: How AI ethics guidelines can be applied and how they can be improved. *AI and Ethics*, 2(1), 2022, 1–13.
- ARAZ TAEIHAGH: Governance of artificial intelligence. A comparative analysis of national strategies. *Policy and Society*, 42(1), 2023, 156–175.
- H. MÜLLER – A. SCHMIDT: Implementing Trustworthy AI. A Pan-European Assessment. *Digital Policy, Regulation and Governance*, 26(3), 2024, 301–320.
- M. KOWALSKI – A. NOWAK: The Right to Explanation in Practice. Challenges and Solutions. *European Data Protection Law Review*, 10(1), 2024, 78–95.
- C. DUBOIS – T. VAN DER MEER: Explainability Requirements Under the EU AI Act. A Technical and Legal Analysis. *AI and Law*, 32(2), 2024, 189–210.