European Journal of Risk Regulation (2023), **14**, 457–475
doi:10.1017/err.2023.61

CAMBRIDGE UNIVERSITY PRESS

SYMPOSIUM ON CLIMATE, AI & QUANTUM

Adjudication of Artificial Intelligence and Automated Decision-Making Cases in Europe and the USA

Elif Kiesow Cortez and Nestor Maslej

Stanford University, Stanford, CA, USA
Corresponding author: Elif Kiesow Cortez; Email: elifkiesowcortez@gmail.com

Abstract

Artificial Intelligence (AI) has started to impact many facets of the economy and people's routine activities. This article contributes to our understanding of how the legal system is reacting to the ongoing uptake of AI and the disputes or right infringements this uptake creates. Select legal cases regarding the use of AI technology for automated decisions are reviewed, with a focus on filings in Europe and the USA. This exercise reveals which type of legal challenges can be expected when it comes to deploying automated systems in these jurisdictions. Additionally, incipient regulatory efforts targeting AI on both sides of the North Atlantic are introduced and briefly discussed. The paper sheds light on how different legal systems accommodate an emerging technology with disruptive potential and offers a mapping of exemplary legal risks for prospective actors or organisations seeking to develop and deploy AI.

Keywords: Adjudication of artificial intelligence; administrative decisions on artificial intelligence; artificial intelligence; artificial intelligence regulation; bias; court cases; privacy

I. Introduction

Artificial Intelligence (AI) technology is becoming more prevalent and is increasingly deployed by economic actors in various domains for informing an ever larger number of decisions and practices at many levels. The ubiquity of digital devices in daily life and across the economy was facilitated by there being a continuous decrease in cost per achieved computing performance, resulting in the production of an immense amount of data.[1] These elements combined provide fertile ground for the dissemination and uptake of AI, with the accumulated effects on economic activity and society overall being difficult to anticipate.[2]

This article offers insights into how rule-of-law institutions are reacting to the increasing deployment of AI systems and, in doing so, provides an idea of the various collective responses to AI adoption, the guardrails being developed to contain abuses and in general the attempt to strike the right balance in achieving benefits while minimising the risks of this powerful technology.[3] This paper will offer a legal analysis of select legal

[1] J Koomey et al, "Implications of Historical Trends in the Electrical Efficiency of Computing" (2011) 33 IEEE Annals of the History of Computing 46.

[2] A Agrawal, JS Gans and A Goldfarb, "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction" (2019) 33 Journal of Economic Perspectives 31.

[3] M-F Cuéllar, "A Common Law for the Age of Artificial Intelligence: Incremental Adjudication, Institutions, and Relational Non-Arbitrariness" (2019) 119 Columbia Law Review 1773.

cases regarding the use of AI technology for automated decisions, oftentimes involving the (mis)use of personal data. The focus lies on exemplary disputes drawn from the US and the EU, and these jurisdictions make for interesting cases in light of the proposed EU Artificial Intelligence Act,[4] the US Algorithmic Accountability Act that was reintroduced in 2022[5] and the US AI Bill of Rights of the same year.[6] This paper aims to provide an initial overview and analysis of emerging regulatory frameworks that affect this breakthrough technology in the USA and the European Union (EU) and will bring to light which type of legal challenges can be expected when it comes to deploying AI-based automated decision-making in these jurisdictions. This paper will also highlight the role of pursuing a risk-based approach in the context of emerging AI technologies.

More specifically, the paper presents an overview and analysis of selected cases from the EU-based judiciary and administrative authorities and from US courts on a range of AI and automated decision-making subjects, including automated web scraping, facial recognition, voice recognition for detecting emotions, government use of digital welfare fraud detection systems, algorithms assisting judicial decisions, automated recommender systems in social networks in connection with harmful content and the use of digital fraud detection systems for social and unemployment benefits. The information from the US and EU case studies offered in this paper will improve our understanding of the legal hurdles facing AI deployment and widen our knowledge on jurisdictional heterogeneity in legal treatment and therefore contribute to reducing legal uncertainty around making AI work in practice. In order to better capture the current status quo judicial reasoning on AI as reflected in the rulings, the article focuses on select cases from the time period between 2017 and 2022.

According to the draft EU Artificial Intelligence Regulation, an AI system is defined as follows: "'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I[7] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with". Throughout the paper, we will be using the definition broadly to cover contemporary automated systems and how the existing case law and enforcement decisions address these systems, with an aim to inform the current and future debates on how AI systems are likely to be treated by the judiciary.

II. Emerging AI rules

This section will cover the central current regulatory frameworks existing or being developed that are applicable to AI and automated decision-making systems in the EU and in the USA.

1. EU Artificial Intelligence Act

The European Commission proposed the Artificial Intelligence (AI) Act in April 2021 and stated that the proposed regulation will serve the goal of enabling the EU to build strategic

[4] "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 21 April 2021, COM2021/0206 – C9-0146/2021 – 2021/0106(COD)" (2021).

[5] S.3572 – Algorithmic Accountability Act of 2022, "S.3572 – 117th Congress (2021–2022): Algorithmic Accountability Act of 2022" (3 February 2022) http://www.congress.gov/ (last accessed 20 December 2022).

[6] "Blueprint for an AI Bill of Rights | OSTP" (*The White House*) https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (last accessed 20 December 2022).

[7] The techniques underlying AI systems as per the EU AI Act are listed under Annex I and include the following categories: "(a) Machine learning approaches, (b) Logic- and knowledge-based approaches, (c) Statistical approaches".

leadership in high-impact sectors, facilitating the development and uptake of AI in the EU, making the EU the place where AI thrives from the lab to the market and ensuring that AI works for people and is a force for good in society.[8]

First, the EU AI Act is set out to be a horizontal regulation, which means that it formulates a common set of rules on AI applicable horizontally across the different sectors of the economy.[9] In contrast, the regulatory framework that might evolve in the USA could be drastically different. For example, although it is difficult to predict at this point, it is conceivable that the regulatory framework for AI and automated systems emerging in the USA will evolve more in the direction of a sector-based approach, at least when compared to the European regime.

Second, the approach taken in the EU AI Act is risk-based in the sense that application domains in which AI will be deployed are categorised according to the risk that is deemed to be posed by utilising AI in that specific domain. The European Commission, as drafter of the Act, expects most uses of AI to pose "low risk" and, following the EU AI Act's risk-based approach, exempts the vast majority of applications of AI from any binding obligations. At the other extreme, the Commission identified and listed a select number of use cases for which the use of AI is forbidden due to "unacceptable risk". Among these completely banned practices are using AI for social scoring or using AI systems that exploit the vulnerabilities of a specific group of persons. The use cases falling within the "high-risk" category might be most interesting to the reader because for this category the EU AI Act requires concrete and binding measures (ie an *ex-ante* conformity assessment before the AI system is put on the market accompanied by an ongoing risk management system for when it has entered the market).

Another aspect of the draft EU AI Act is its extraterritorial application, which means that it has the potential to influence practices beyond Europe's borders. Regarding its sanctioning regime, the draft Act foresees three levels of severity of fines. The most significant fines are levied for violating the prohibition of specific AI systems, and they can go up to €30 million or 6% of the violating company's turnover. A considerable part of the fine regime concerns the rules for high-risk systems, and it targets the different actors along an AI system's life cycle (ie not only "providers" but also "users",[10] importers, distributors and notified bodies can be subject to high sanctions). However, the draft Act also determines that small and medium-sized enterprises (SMEs) and start-ups would face relatively lower fines for infringements.

2. *US Algorithmic Accountability Act 2022*

In February 2022, the Algorithmic Accountability Act of 2022 (AAA) was presented to the US Congress.[11] In light of the unease associated with the use of automated systems in reaching decisions in domains such as housing, education, employment, healthcare and lending, the AAA is supposed to enhance the transparency of algorithm deployment in various contexts in order to minimise discriminatory, biased or harmful decisions.

In general, the goal of the AAA is to be able to hold organisations to account when they deploy algorithms or automated systems for generating decisions in a manner that

[8] "A European Approach to Artificial Intelligence | Shaping Europe's Digital Future" https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence (last accessed 19 December 2022).

[9] This stands in contrast to, for instance, one of the discarded alternative regulatory approaches that was considered, an approach that would consist of various sector-specific rules and obligations.

[10] For clarification, "users" as defined in the EU AI Act are basically organisations deploying the AI system.

[11] An earlier version of this Act was introduced in April 2019 but did not garner sufficient backing for enaction. Major revisions were made in the 2022 version – for instance, when compared to the earlier 2019 version, the new version contains stricter impact assessment requirements. However, it is also still unclear whether or when the 2022 version could become law.

significantly affects individuals in the USA. In this line, the AAA proposes government-mandated "impact assessments" for organisations that use automated decision systems, and the impact assessments follow a two-pronged approach as they should be conducted both before deployment and after deployment in the form of augmented decision-making processes.[12]

Many commenters have pointed out some of the AAA's shortcomings. First, it has been noted that the Act applies only to large private companies, exempting SMEs and government agencies from regulatory scrutiny.[13] Second, the AAA is blamed for at times lacking detail and specificity. For one, a number of relevant policy choices are left to be determined by the Federal Trade Commission. In addition, the language is criticised for being vague at times – for example, relying on qualifiers such as "to the extent possible", which can potentially lead to enforcement challenges.[14] Furthermore, commenters have criticised the AAA for not taking into account some potential trade-offs involved – for instance, with regard to safeguarding equal treatment of protected groups while simultaneously aiming for accuracy, efficiency and privacy.

It is not immediately clear how strong of a priority it will be for the US administration to get the AAA passed in the near future, with approval pending in the House of Representatives and the Senate. However, given the federal structure of the US government, regulation of AI systems can already start at the subnational level, as exemplified by New York City's Law on Automated Employment Decision Tools, which requires that automated systems used for hiring decisions from January 2023 onwards get audited for bias by an independent auditor that assesses potential disparate impacts on certain groups.[15] On 12 December 2022, the Department of Consumer and Worker Protection (DCWP) announced that it would postpone enforcement until 15 April 2023.[16] According to the agency's statement, a second public hearing is being planned due to the high volume of public comments.[17]

3. *US AI Bill of Rights*

The White House's "blueprint" for an AI Bill of Rights (AIBoR) was published in October 2022. It does not constitute hard law; however, it provides a relevant framework that will potentially guide and inform AI laws and policymaking in the USA. This is the latest and thus far most evident initiative undertaken by the Biden administration to regulate AI and algorithms. The AIBoR lays out AI-related potential civil rights harms and presents five principles to be observed as well as providing a more detailed companion containing guidance on how to implement the principles.

The first principle demands AI systems to be "safe and effective", which calls for pre-deployment testing for risks and harm mitigation. Second, "notice and explanation"

[12] J Mökander et al, "The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?" (2022) 32 Minds and Machines 751.

[13] J Mökander and L Floridi, "From Algorithmic Accountability to Digital Governance" (2022) 4 Nature Machine Intelligence 508.

[14] ibid.

[15] BC Larsen, "The Geopolitics of AI and the Rise of Digital Sovereignty" (*Brookings*, 8 December 2022) https://www.brookings.edu/research/the-geopolitics-of-ai-and-the-rise-of-digital-sovereignty/ (last accessed 14 December 2022).

[16] "AI Bias Law Postponed until April 15 as Unanswered Questions Remain" (*VentureBeat*, 12 December 2022) https://venturebeat.com/ai/for-nycs-new-ai-bias-law-unanswered-questions-remain/ (last accessed 19 December 2022).

[17] "New Laws & Rules – DCA" https://www.nyc.gov/site/dca/about/new-laws-rules.page#:~:text=December%202022%20Update%3A%20The%20Department,144%20until%20April%2015%2C%202023 (last accessed 19 December 2022).

should be provided, a principle that emphasises both making persons aware that they are facing an AI system and being to some degree transparent by way of providing information on how the AI system generates decisions. Third, a "right to data privacy" should be observed – this principle relates to persons retaining control over how their data are used and prescribes data minimisation, but it also calls for heightened oversight for AI in the context of surveillance systems. Fourth, the right to "protection from algorithmic discrimination" demands assessments regarding equity and continuous alertness to and mitigation of disparities. The fifth and last principle asks for "human alternatives, consideration, and fallback", providing persons with a way to opt out or to access a human with the means to review and override decisions generated by algorithm.

The AIBoR focuses on AI systems used in human services – for example, in education, health services, lending, hiring, and commercial surveillance, among others. Hence, it is not a comprehensive AI guidance, as it lacks emphasis on the utilisation of algorithms within most consumer products, online information ecosystems or critical infrastructure.[18]

4. EU–US roadmap

In order to promote responsible AI on both sides of the North Atlantic, the USA and the EU published a "joint roadmap on evaluation and measurement tools for trustworthy AI and risk management" on December 2022.[19] The roadmap is issued as an output of the EU–US Trade and Technology Council, and it states as one of its aims the commitment to the Organisation for Economic Co-operation and Development's (OECD) recommendations on AI.[20]

In May 2022, the US–EU Joint Statement of the Trade and Technology Council declared that that the parties confirm their "commitment to collaboration in developing and implementing trustworthy AI through a human-centered approach that reinforces shared democratic values and respects human rights".[21] The report also declared that intentions exist to develop a joint roadmap on AI risk management. It was stated that the parties are dedicated "to develop[ing] a shared hub/repository of metrics and methodologies for measuring AI trustworthiness and AI risks" so that this repository could be used to evaluate the technical requirements for trustworthy AI and bias mitigation as well as facilitating interoperable approaches to managing AI risks.[22]

In December 2022, the EU–US "Joint Roadmap for Trustworthy AI and Risk Management" was made public, with the aim of advancing shared terminologies and taxonomies and fomenting transatlantic communication on metrics for measuring AI trustworthiness and risk management methods.[23] In this report, both parties reaffirmed that "a risk-based approach and a focus on trustworthy AI systems can provide people with

[18] A Engler, "The AI Bill of Rights Makes Uneven Progress on Algorithmic Protections" (*Brookings*, 21 November 2022) https://www.brookings.edu/2022/11/21/the-ai-bill-of-rights-makes-uneven-progress-on-algorithmic-protections/ (last accessed 14 December 2022).

[19] EU–US TTC, "TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management" https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management (last accessed 14 December 2022).

[20] OECD AI Principles overview https://oecd.ai/en/ai-principles (last accessed 19 December 2022).

[21] US–EU Trade and Technology Council, "U.S.–EU Joint Statement of the Trade and Technology Council" (2022) https://www.commerce.gov/sites/default/files/2022-05/US-EU-Joint-Statement-Trade-Technology-Council.pdf (last accessed 18 December 2022).

[22] ibid.

[23] "TTC Joint Roadmap for Trustworthy AI and Risk Management | Shaping Europe's Digital Future" https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management (last accessed 19 December 2022).

confidence in AI-based solutions, while inspiring enterprises to develop trustworthy AI technologies".[24] The report emphasises the importance of using a risk-based approach to AI, which may be the key to developing trustworthy AI systems that enhance innovation, reduce trade barriers, boost market competition, operationalise common values and protect human rights.

III. Case law selection: a transatlantic perspective

When a revolutionary technological innovation starts being utilised by economic actors, the legal systems begin to process its potential ramifications and offer a legal response. We will be providing an overview and analysis of selected legal cases that concern AI: three cases from EU judiciary and administrative authorities and three cases from US courts from the time period ranging from 2017 until 2022 in order to better understand how actors in the legal system in different jurisdictions perceive the risks and benefits of the use of AI and how these perceptions currently translate to judicial and administrative decisions.

The selection of cases was done by taking into consideration the risk-level classification for AI in the respective use case domains connected to each particular adjudication example. The chosen cases would probably fall under the defined high-risk category within the draft EU AI Act, which is expected to cover the most relevant type or class of cases for practitioners, policymakers and judiciary officials given that for the high-risk category binding and enforceable provisions will be applicable. Given the limited scope of the article, three cases are selected from the EU and the USA each to cover AI in the realm of public safety, AI in the context of fraud detection for welfare provision and the use of AI by private actors for recognition and recommender systems.

It is, then, of the greatest relevance to enhance our understanding of how the judiciaries and administrative systems are already in the present issuing decisions on cases that might be in the future classified or considered as high risk. The selection of the cases was also made in order to be helpful in the sense that areas or legal issues that will not be covered by future AI regulatory regimes will be informed by the already-ruled-on cases from the past and the principles that emerged in their adjudication. This can be especially relevant for the USA, where one does not know whether AI policy will be likely to rely on a "soft law"-oriented approach or to be determined individually by (select) federal subunits, and where the judiciary would take into account and apply the current body of law to AI-related disputes.

In Europe, the digital policy regime is in a state of flux, with multiple initiatives targeting the digital economy preoccupying European lawmakers. While these EU-level laws slowly take shape, get passed and, years later, enter into effect, it is relevant to provide information on examples of how AI systems are being handled already in the present day. Some of the current decisions and emerging principles might shape how some lesser-defined aspects of future laws might ultimately be interpreted.

1. Case law examples: Europe

The following section presents selected cases from the EU that involve disputes on issues such automated web scraping, facial recognition, voice recognition for detecting emotions and government use of digital welfare fraud detection systems.

[24] EU–US TTC, supra, note 19.

a. Childcare benefits and SyRI cases: The Netherlands (25 November 2021 and 5 February 2020)

A December 2022 report on the bias in algorithms, AI and potentially discriminatory outcomes prepared by the EU Fundamental Rights Agency states in its foreword that AI-based algorithms are becoming increasingly important as they are used for essential decision-making systems such as "determining who will receive state benefits".[25] In 2021, the Dutch government resigned due to what was called the "childcare scandal", as it was discovered that the algorithm used for risk assessment to detect welfare fraud was biased towards targeting foreigners.[26] An Amnesty International report pointed out that the algorithm reinforced the existing institutional bias of a racial/ethnic link between crime and race and ethnicity and was deployed with no meaningful human oversight. The self-learning mechanism reproduced discriminatory design flaws by adapting the algorithm based on experience over time.[27] The report described this process as a discriminatory loop in which non-Dutch nationals were flagged more frequently for fraud than Dutch nationals.[28]

On 25 November 2021, the Dutch Data Protection Authority fined the Minister of Finance €2.75 million, as the Tax and Customs Administration automatically categorised the risk of certain applications for benefits while making use of an algorithm that used the nationality of applicants as an indicator of risk (Dutch/non-Dutch).[29] Using a self-learning algorithm, the risk classification model tested all applications within a given month, and this algorithm automatically selected requests for which available human resources were deployed. In that month, based on the algorithm's risk assessment, the 100 applications with the highest risk score were presented to employees for manual review. It was noticed that from March 2016 to October 2018 the model included an indicator "Dutch nationality/non-Dutch nationality".

Amnesty International reported that the Dutch government had knowledge of the risk of the human rights impacts of using algorithmic decision-making systems such as the digital welfare fraud detection system SyRI (*systeem risico indicatie*; system risk indication).[30] The Hague District Court ruled on 5 February 2020 that the Dutch government can no longer use SyRI on the basis that it violated the European Convention on Human Rights (ECHR) Article 8, which protects the right to respect for private and family life.[31] SyRI was a system used for its ability to assist the government in identifying

[25] EFRA, *Bias in Algorithms: Artificial Intelligence and Discrimination* (Publications Office, 2022) https://data.europa.eu/doi/10.2811/25847 (last accessed 20 December 2022); S Wachter, "Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR" (2018) 34 Computer Law & Security Review 436.

[26] S Amaro, "Dutch Government Resigns after Childcare Benefits Scandal" (*CNBC*, 15 January 2021) https://www.cnbc.com/2021/01/15/dutch-government-resigns-after-childcare-benefits-scandal-.html (last accessed 20 December 2022); "Dutch Childcare Benefit Scandal an Urgent Wake-Up Call to Ban Racist Algorithms" (*Amnesty International*, 25 October 2021) https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/ (last accessed 20 December 2022).

[27] Amnesty International, "Xenophobic Machines: Discrimination Through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal" (2021) https://www.amnesty.org/en/documents/eur35/4686/2021/en/ (last accessed 14 December 2022).

[28] ibid.

[29] "Boete Belastingdienst voor discriminerende en onrechtmatige werkwijze" https://autoriteitpersoonsgegevens.nl/nl/nieuws/boete-belastingdienst-voor-discriminerende-en-onrechtmatige-werkwijze (last accessed 20 December 2022). The decision can be accessed at the Dutch Data Protection Authority website: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/boetebesluit_belastingdienst.pdf (last accessed 19 December 2022).

[30] Amnesty International, supra, note 27.

[31] "SyRI legislation in breach of European Convention on Human Rights" https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx (last accessed 20 December 2022). The decision can be accessed at the Hungarian Data Protection Authority website: https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:865 (last accessed 19 December 2022).

individuals at risk of engaging in fraud when it comes to social security, taxation and employment laws.[32]

The case also holds special importance as it is one of the world's first court decisions to stop the use of digital welfare technologies on human rights grounds as observed by the UN Special Rapporteur on extreme poverty and human rights.[33] In its analysis of the SyRI case, the court expressed similar concerns as the Special Rapporteur and warned that SyRI could discriminate on the basis of socio-economic status and migrant status without additional information regarding how the automated system works.[34]

The decision of the court states that Dutch Section of the International Commission of Jurists (NJCM) et al,[35] who were being joined in this procedure by the Dutch Trade Union Confederation (FNV) and the Special Rapporteur on extreme poverty and human rights, have provided detailed explanations of their view that SyRI is discriminatory and stigmatising. It was argued by the parties that SyRI would facilitate further investigation of neighbourhoods that were currently recognised as problematic areas. Thus, they suggested that there would be a greater likelihood that irregularities would be found in those than in other neighbourhoods. It was argued that this possibility would reinforce a negative perception of the relevant neighbourhoods and of their residents and strengthen stereotyping even though no risk notifications had been issued.[36]

In the aftermath of the case, it was announced that as of 1 January 2023, the Dutch Data Protection Authority would play a special role in algorithm supervision. The letter by the State Secretary of Kingdom Relations and Digitalisation announcing this new role also emphasises the importance of supervising deployed algorithms to prevent repetition of the childcare scandal in the future.[37]

The significant impacts that can result from government institutions deploying automated decision-making in sensitive areas, with a very large number of citizens potentially affected by biases or discriminatory algorithms, call for carefully designed rules and thorough assessment when relying on such systems in the public sector. The last case that will be discussed concerning the US context (Section III.2.c) is about a digital fraud detection system ("MiDAS") utilised by authorities in Michigan to automatically identify fraudulent behaviour or discrepancies in the receipt of unemployment benefits. It is a very rudimentary tool operated by way of using logic trees, meaning that it is much simpler than more advanced AI systems of the type currently being released. This case connects in some respects to the case of the flawed detection algorithm SyRI deployed in the Netherlands to scan for social benefit fraud. Whereas the Dutch SyRI case centres on algorithm bias, in the case of MiDAS in Michigan the tool's error-proneness was the central issue. The core of the plaintiffs' complaints regarding MiDAS is that using the automated system resulted in them being deprived of unemployment benefits based on non-transparent procedures and without adequate pre- or post-deprivation processes. This case also indicated that liability could be established for private entities that assist in the development of AI-like tools deployed, for instance, by public authorities, and that a

[32] EFRA, supra, note 25.

[33] "Landmark Ruling by Dutch Court Stops Government Attempts to Spy on the Poor – UN Expert" (*OHCHR*, 5 February 2020) https://www.ohchr.org/en/press-releases/2020/02/landmark-ruling-dutch-court-stops-government-attempts-spy-poor-un-expert (last accessed 20 December 2022).

[34] ibid.

[35] These proceedings were initiated against the Netherlands by a number of civil society interest groups, including the NJCM and two private individuals. The FNV joined as a party to the claimants' proceedings.

[36] Rechtbank Den Haag, "ECLI:NL:RBDHA:2020:865" (2020) https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:RBDHA:2020:865 (last accessed 19 December 2022).

[37] Inrichtingsnota algoritmetoezichthouder, 22 December 2022, 26643–953, available at: https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022Z26092&did=2022D56234 (last accessed 06 January 2023).

company might become legally liable even though it did not directly oversee the tool's actual use.

b. Budapest Bank case: Hungary (8 February 2022)

Budapest Bank has been fined €634,000 by the Hungarian Data Protection Authority (NAIH) on the basis that the bank automated the evaluation of customers' emotional states using an AI-driven software solution.[38] Using the speech evaluation system, the bank determined which customers needed to be attended to based on their mood. This application was used to prevent complaints as well as to retain customers.[39]

Reporting the phone calls of customers is common practice among many companies. In a former case brought to the attention of the French Data Protection Authority (CNIL), recording customer phone calls only led to a fine when other General Data Protection Regulation (GDPR) requirements such as implementing data minimisation were not followed.[40] In other words, solely the recording of customer calls is not prohibited under the GDPR as long as the other requirements of the regulation are complied with. However, with the Budapest Bank case, in addition to recording all customer service telephone calls, the data controller automatically analysed all new audio recordings every night using AI. The software aimed to predict the emotional state of the client at the time of the call based on the keywords that were detected. In conjunction with the voice call, the emotional analysis results of the data subjects were also stored within the system for forty-five days.[41]

The key findings of the decision reported by the European Data Protection Board (EDPB) states that, based on the audio recording of the customer service telephone calls, a list of customers was compiled that assessed their likelihood of being dissatisfied and angry. The results of the analysis were utilised to identify clients whose dissatisfaction could be addressed by customer service personnel. The data subjects were not informed about this particular data processing, and they were not technically entitled to object, yet the data processing was planned and carried out nevertheless. Additionally, the data controller's impact assessment confirmed that the reviewed data processing relied on AI and posed a high risk to the fundamental rights of individuals.[42]

The decision text details the NAIH's analysis of the impact assessment, demonstrating that the assessment was deemed by the authority as formally correct; however, its content did not correspond to reality. It was concluded that the assessment did not deal with the issue of the analysis of emotions in any meaningful way,[43] and based on the company's statements it was decided that the company was clearly aware of these shortcomings when preparing the impact assessment, as well as during the mandatory regular checks during operational and GDPR compliance reviews.[44]

[38] "Data Protection Issues Arising in Connection with the Use of Artificial Intelligence | European Data Protection Board" https://edpb.europa.eu/news/national-news/2022/data-protection-issues-arising-connection-use-artificial-intelligence_en (last accessed 19 December 2022). The decision can be accessed at the Hungarian Data Protection Authority website: https://www.naih.hu/hatarozatok-vegzesek?download=517:mesterseges-intelligencia-alkalmazasanak-adatvedelmi-kerdesei (last accessed 19 December 2022).

[39] ibid.

[40] CNIL, "Délibération SAN-2020-003 Du 28 Juillet 2020" (2020) https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000042203965/ (last accessed 14 December 2022).

[41] "Data Protection Issues Arising in Connection with the Use of Artificial Intelligence | European Data Protection Board", supra, note 38.

[42] ibid.

[43] P Lewinski, J Trzaskowski and J Luzak, "Face and Emotion Recognition on Commercial Property under EU Data Protection Law" (2016) 33 Psychology & Marketing 729.

[44] NAIH, "Case Number: NAIH-85-3/2022" (2022) https://www.naih.hu/hatarozatok-vegzesek?download=517:mesterseges-intelligencia-alkalmazasanak-adatvedelmi-kerdesei (last accessed 14 December 2022).

The key findings summarised by the EDPB report state that it was clear from both the impact assessment and the legitimate interest assessment that no actual risk mitigation measures were provided, and only some insufficient measures such as information and right of objection existed, and these were merely applicable in theory. AI systems are complex. Moreover, these systems are often prone to bias and do not readily enable the verification of the data they process. Therefore, the transparent and safe deployment of AI tools requires not only additional safeguards, but also ones that are also clearly delineated.[45]

The proposed EU AI Act[46] addresses the issue of emotion recognition systems by defining these systems in Article 3.34 as "an AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data". A report by the European Parliament's Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs published on April 2022 cites the amendments suggested by the Parliament regarding the definition of emotion recognition systems, and these amendments propose to expand the relevant definition in the following manner "... inferring emotions, thoughts, states of mind or intentions of natural persons ...".[47]

The explanatory memorandum of the EU AI Act attends to this issue by indicating that the transparency obligations will be applicable to AI systems "that (i) interact with humans, (ii) are used to detect emotions or determine association with (social) categories based on biometric data, or (iii) generate or manipulate content ('deep fakes')". The relevant section of the EU AI Act also states that individuals must be informed when they interact with AI systems or when their emotions or characteristics are being recognised through automated means, and Recital 70 of the current version of the proposed Act reads "... natural persons should be notified when they are exposed to an emotion recognition system or a biometric categorization system".[48]

The notion of utilising AI for the purpose of emotion recognition preoccupies legislators, especially in Europe. Doing so would open up opportunities for market actors to engage in profit-seeking manipulation – for instance, via targeted advertisement or customer treatment.

c. Clearview AI case: France, Italy and Greece (17 October 2022)

The last case in this section covering Europe-based filings will focus on the decision regarding Clearview AI, a software company that extracted high volumes of publicly available images. On 17 October 2022, CNIL issued a penalty of €20 million against Clearview AI. This was the third fine of the same amount issued against the company in 2022 by EU data protection authorities. The decision was aligned with the earlier decision of the Italian Data Protection Authority, which issued another €20 million fine against Clearview AI on 10 February 2022. The company also faced the same level of fine by the

[45] "Data Protection Issues Arising in Connection with the Use of Artificial Intelligence | European Data Protection Board", supra, note 38.

[46] "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 21 April 2021, COM2021/0206 – C9-0146/2021 – 2021/0106(COD)", supra, note 4.

[47] Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs, "DRAFT REPORT on the Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM2021/0206 – C9 0146/2021 – 2021/0106(COD))" (2022).

[48] "Proposal for a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 21 April 2021, COM2021/0206 - C9-0146/2021 - 2021/0106(COD)", supra, note 4.

Greek Data Protection Authority on 13 July of the same year.[49] In addition, on 23 May 2022, the company was fined £7.5 million by the UK Data Protection Authority (Information Commissioner's Office; ICO) and was ordered to delete the data belonging to UK residents.[50]

Clearview AI extracts and collects images, photos and videos from a variety of sources, including social media. More specifically, all of the photos that the company collects are directly accessible without creating an account or logging in.[51] On its own website, Clearview AI promotes their product as an "investigative platform" that "allows law enforcement to rapidly generate leads to help identify suspects, witnesses and victims".[52] The company reportedly has amassed over 20 billion publicly available facial images.[53] Given this astonishingly large dataset, the company sells to law enforcement agencies access to its database and promotes an image search engine to facilitate the identification of suspects and victims.[54]

In their analysis of the case, the CNIL decision states that an extensive amount of photographic data was collected by the company in the present case,[55] which is associated with other personal data likely to reveal a variety of aspects of a person's private life, which constitutes an extremely intrusive form of processing. These data were then used to create a biometric template (ie biometric data), assuming they were reliable, allowing for the unique identification of a person from a photograph of the individual. The authority declares that the detention of such data by a third party also constitutes a significant privacy invasion.

CNIL also refers to the reasonable expectation of individuals regarding the privacy of their photos. The ruling highlights that it is reasonable for the data subjects to expect that third parties will be able to access the photographs from time to time; however, the public nature of these photographs cannot be considered sufficient to conclude that the subjects can reasonably expect their images to be used in data processing. It is also pointed out in this justification that the software used by the company is not publicly available, and the vast majority of those concerned are unaware of its existence. This section of CNIL's decision may be used to draw parallels to a common US privacy concept: the reasonable expectation of privacy. In the well-known decision Katz v. United States, 389 U.S. 347 (1967), a two-pronged test was developed to assess whether the individual's rights under the Fourth Amendment were infringed. The test looked at whether the individual's

[49] "Facial Recognition: Italian SA Fines Clearview AI EUR 20 Million | European Data Protection Board" https://edpb.europa.eu/news/national-news/2022/facial-recognition-italian-sa-fines-clearview-ai-eur-20-million_en (last accessed 19 December 2022). The decision can be accessed at the Greek Data Protection Authority website: https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9751362 (last accessed 19 December 2022); "Hellenic DPA Fines Clearview AI 20 Million Euros | European Data Protection Board" https://edpb.europa.eu/news/national-news/2022/hellenic-dpa-fines-clearview-ai-20-million-euros_en (last accessed 19 December 2022). The decision can be accessed at the Greek Data Protection Authority website: https://www.dpa.gr/el/enimerwtiko/prakseisArxis/epiboli-prostimoy-stin-etaireia-clearview-ai-inc (last accessed 19 December 2022).

[50] "Clearview AI Inc." (26 May 2022) https://ico.org.uk/action-weve-taken/enforcement/clearview-ai-inc-mpn/ (last accessed 19 December 2022).

[51] "Facial Recognition: 20 Million Euros Penalty against CLEARVIEW AI | CNIL" https://www.cnil.fr/en/facial-recognition-20-million-euros-penalty-against-clearview-ai (last accessed 19 December 2022).

[52] "Clearview AI | Facial Recognition" (*Clearview AI*) https://www.clearview.ai (last accessed 19 December 2022).

[53] "Clearview AI Now Features 20B Facial Images" https://iapp.org/news/a/clearview-ai-now-features-20-billion-facial-images/ (last accessed 19 December 2022).

[54] "Facial Recognition: 20 Million Euros Penalty against CLEARVIEW AI | CNIL", supra, note 51; see also: EK Cortez, "Data Protection Around the World: Future Challenges" in EK Cortez (ed.), *Data Protection Around the World: Privacy Laws in Action* (The Hague, Springer – TMC Asser Press 2021); K Ringrose, "Law Enforcement's Pairing of Facial Recognition Technology with Body-Worn Cameras Escalates Privacy Concerns Essay" (2019) 105 Virginia Law Review Online 57; Y Welinder, "A Face Tells More than a Thousand Posts: Developing Face Recognition Privacy in Social Networks" (2012) 26 Harvard Journal of Law & Technology 165.

[55] Reportedly 20 billion images; see " Clearview AI Now Features 20B Facial Images", supra, note 53.

subjective expectation of privacy was violated and whether this expectation of privacy was reasonable. CNIL declares that the mere fact that an individual made their photo public does not override their reasonable expectation that this photo will not be used by a company to create a biometric profile of that individual.

In fact, CNIL's decision is aligned with a recent report of an individual who filed a complaint with his local data protection authority in Germany against Clearview AI. Matthias Marx reported that he filed this complaint as he found out that his face had been mapped and monetised by three companies.[56] According to the article, Marx states that he read about Clearview AI in a report stating that it scraped billions of photos from the Internet to create a huge database of faces in 2020. With Clearview AI's facial recognition technology, law enforcement agencies can find other online photos of the same face by uploading a single photo.[57] He emailed Clearview AI to see whether the company had any photos of his face. In response, Marx received two screenshots from a Google engineering competition that he had attended around a decade prior. He reports that he knew that the pictures existed, but that he did not know that a photographer was selling them without his permission on the stock photo site Alamy.[58]

The CNIL decision focused on posting a photo publicly and how this act does not mean that the user should expect that it will be processed by a large software company serving law enforcement's needs. However, Marx's story shows evidence that it might even be the case that these images might also be used by other third parties for commercial profit.

The CNIL decision was taken under the GDPR, which has been designed by the EU as a regulation with extraterritoriality. However, it can be quite difficult to operationalise extraterritoriality given the limitations of national jurisdictions and international enforcement. As reported by Meaker, at the time when the article was written in November 2022, Clearview AI had not deleted the photographs belonging to EU data subjects, and it was stated that the fines by the Italian and Greek data protection authorities had not been paid, and CNIL did not disclose whether its fine had been paid.[59] This might demonstrate the importance of the potential benefits of further transatlantic alignment via the EU–US common roadmap for assessing AI risks and interoperable approaches to managing these risks, as covered in Section II.4.

2. *Case law examples: USA*

The selected US cases cover a range of subjects, including algorithms assisting judicial decisions, automated recommender systems in a social network and harmful content and the use of digital fraud detection systems for unemployment benefits.

a. Flores v. Stanford (28 September 2021)

This case concerned an application made by Northpointe, Inc., to prevent the disclosure of materials produced by Northpointe to one of the plaintiff's experts, Dr Rudin. Ultimately, Northpointe's request was denied and the compelled materials were ordered to be released under a supplemental protective order. The court ruled that the compelled materials were relevant to the plaintiff's constitutional claims, carried little risk of competitive injury and, if not admitted, could result in an elevated risk of prejudice to plaintiffs.

This particular decision stemmed from a previous court order, dated 12 February 2021, which ordered Northpointe to produce for the plaintiffs a variety of proprietary

[56] M Meaker, "Clearview Stole My Face and the EU Can't Do Anything About It" (*Wired*, 2022) https://www.wired.com/story/clearview-face-search-engine-gdpr/ (last accessed 19 December 2022).

[57] ibid.

[58] ibid.

[59] ibid.

information on the topic of Northpointe's Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool. More specifically, COMPAS is a secret algorithm that was used by the defendants (more broadly the New York State Board of Parole) to decide whether to grant offenders that were sentenced to life in prison discretionary parole. In the former dispute, the plaintiffs contended that the defendants' use of an algorithm that they did not fully understand took away the individualised parole assessment that juvenile offenders sentenced to life are entitled to under the 8th and 14th amendments.

In the previous decision, the court ordered that Northpointe release a variety of proprietary data on the operation of their algorithm to the plaintiffs. However, in order to respect Northpointe's concern that the compelled materials contained trade secrets and other proprietary info, the court mandated that the compelled materials be released under a protective order. Although Northpointe and the plaintiffs entered into a Second Supplemental Stipulation of Confidentiality and Proposed Protective Order on 26 February 2021, Northpointe would not agree to permitting Dr Rudin, one of the plaintiff's expert witnesses, access to the compelled materials. Northpointe contended that the compelled materials contain highly proprietary information and that releasing this information to Dr Rudin, a highly outspoken critic of Northpointe, would jeopardise Northpointe's existence.

The court noted as well that there is no absolute privilege against disclosing trade secrets.[60] In such matters, the court must balance the relevance of commercially sensitive materials to a party's claim (in this case having its chosen expert review the materials) against the interest of the opposing party in maintaining trade secrets.[61] The relevant issues that figure in this consideration are: (1) whether the person receiving the confidential information is involved in competitive decision-making relating to the subject of the materials; (2) whether there is a risk of inadvertent disclosure of the information; (3) the hardship imposed by the restriction; (4) the time of the remedy; and (5) the scope of the remedy.[62]

The court ruled that the disclosure of the compelled materials was relevant, posed little risk of competitive injury and would prevent prejudice against the plaintiff. First, the defendants claimed that the compelled materials were irrelevant to the issue of class certification and only minimally relevant to the lawsuit in its entirety, because only one of the named plaintiffs was denied parole as a result of a high COMPAS risk score. The plaintiffs conversely contended that the compelled materials were highly relevant to class certification and merits as they showed how the algorithm factors youth into its risk scores across the proposed class and for the individual plaintiff. The court contended that the class issue does in factor overlap quite significantly with merits discovery[63] and held the issue to be relevant given that courts are typically unwilling to separate class-related discovery from discovery of the merits.[64] Moreover, the court said that the compelled materials were highly relevant to the plaintiff's constitutional claims. The plaintiffs in their previous motions contended that COMPAS treats youth as an aggravating factor in their model, which fails to account for juvenile rehabilitative capabilities and ultimately violates the right to individualised parole assessment implied in the 8th and 14th amendments. Northpointe suggested that age is a datapoint, but it has failed to reveal how significantly the algorithm weights age. Therefore, the court ruled that more information

[60] Federal Open Mkt. Comm. of Fed. Reserve Sys. v. Merrill, 443 U.S. 340, 362, 99 S. Ct. 2800, 61 L. Ed. 2d 587 (1987).

[61] Grand River Enters. Six Nations, Ltd v. King, No. 02 Civ. 5068 (JFK), 2009 U.S. Dist LEXIS 11504, 2009 WL 222160, at *5 (S.D.N.Y Jan. 30, 2009 also Uniroyal Chem. Co. Inc. v. Syngenta Crop prot., 224 F.R.D 53, 57 (D. Conn. 2004).

[62] Uniroyal Chem. Corp., 224 F.R.D at 57.

[63] Bodner v. Paribas, 202 F.R.D. 370, 373 (E.D.N.Y. 2000).

[64] Chen-Oster v. Goldman, Sachs & Co., 285 F.R.D. 294, 299 (S.D.N.Y. 2012).

was needed to understand how the algorithm weights age of offense. In addition, this information needs to be adjudicated by an expert witness such as Dr Rudin, a witness who would understand how the algorithm operates.

Second, the court, while sensitive to Northpointe's concern about the dissemination of trade secrets, felt that the risk of intentional or inadvertent disclosure of this information to competitors is low. The court reasoned that the protective orders already in place prevent the disclosure of the compelled materials except for individuals involved in issues relating to the litigation of the case. Moreover, Dr Rudin has agreed to be bound by the orders, has promised to take further measures to prevent inadvertent disclosure and has a history of lawfully dealing with confidential information in the past.

This case seemed relevant for inclusion because it illustrates a tension between algorithm-deploying companies that aim to maintain their trade secrets and opposing parties that want more clarity on how algorithms operate. It is highly likely that these kinds of tensions will increase in the near future as AI systems improve in their capability and are increasingly likely to be deployed in the real world. Companies that use these systems will desire to keep their operation secret for competitive reasons, while individuals that are unfairly affected by these systems will attempt to understand how they operate. This case is illustrative because it paints a picture of the particular legal issues that courts consider in resolving these tensions. In such disputes, the two most salient legal issues seem to be the relevance of proprietary information to the issue at hand and the possibility of competitive injury. The decision in this case suggests that should an algorithm make a decision that affects an individual in a particularly unlawful way, it is acceptable that information on the workings of such algorithms be released in a court setting for the resolution of disputes. A more tricky question concerns the operation of more complex neural network-based algorithms that learn weights that themselves are sometimes not completely understood by the algorithm developers. These algorithms can sometimes be "black boxes", so the question of how their operation relates to law might be challenging to judge.[65] The second issue on risk of competitive injury seems to be taken seriously by courts. The court acknowledged the right companies have to keep information proprietary, but seemed to believe that the risk of disclosure can be prevented with protective orders.

Another interesting issue is raised by this case: namely, the intersection of company rights over proprietary information regarding the design of their AI systems and the demand for enhancing the explainability of AI systems, especially when their outputs affect natural persons rather directly. Providing explainability is not straightforward in every case, especially as AI systems become more complex and rely on deep neural networks and extensive unsupervised learning. Having said that, explainability is clearly heralded as a desirable feature of AI systems, and computer scientists and engineers are developing methods to better trace back and make sense of AI system outputs. Most importantly, emerging legal and policy frameworks are paying attention to the issue – for example, when the US AIBoR puts forward the already-mentioned "notice and explanation" principle, which also emphasises transparency and disclosure of information on how an AI system generated outputs or decisions.

Connecting to the particular case analysed here, this might become a widespread and contentious issue if achieving some degree of explainability is required but AI system proprietors continue to have concerns about granting access to their algorithms because of fears of exposing their trade secrets.

[65] C Dwork and M Minow, "Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law" (2022) 151 Daedalus 309.

b. Dyroff v. Ultimate Software Grp., Inc. (26 November 2017)

This case pits the plaintiff, Kristanalea Dyroff, against the defendant, Ultimate Software. The plaintiff sued Ultimate Software after her adult son, Wesley Greer, died from an overdose of fentanyl-laced heroin bought from a drug dealer that he had met through Ultimate Software's now discontinued social networking website, Experience Project.

The plaintiff advanced seven charges: (1) negligence, (2) wrongful death, (3) premise liability, (4) failure to warn, (5) civil conspiracy, (6) unjust enrichment and (7) violation of the Drug Dealer Liability Act. The crux of the plaintiff's claim is that Ultimate Software is responsible because it mines data from its users' activity on Experience Project, deploys algorithms to glean insights from their posts and uses those insights to make actionable recommendations for users, which in this case channelled her son towards the heroin-related discussion groups that enabled the purchase of tainted drugs. For all claims advanced by the plaintiff, with the exception of the fourth, Ultimate Software asserted immunity under the Communications Decency Act (CDA), 47 U.S.C. § 230(c)(1). This section of the Act allows websites to have immunity regarding third-party content on their platforms unless they are wholly or partially responsible for the creation and development of such content. The plaintiff responded to Ultimate Software's immunity claim by contending that websites need not co-author posts to technically develop content. More specifically, she asserted that content creation can mean materially manipulating such content – for example, guiding the content's generation or generating novel insights from data and using such insights to guide and signal to users. The court ruled that the defendants did in fact have immunity for all claims with the exception of the fourth under the CDA given that they did not develop the content that led the plaintiff's son to purchase the drug. The court likewise held that the website is unlike a brick-and-mortar business that has a duty to warn, and therefore was not obligated to warn Mr Greer of fentanyl-laced drugs.

To begin, it is useful to consider some background on how Ultimate Software's Experience Project website actually worked. On this particular social media platform, users could register with anonymous usernames and join or start groups related to their interests, such as dogs or drugs. Ultimate Software then deployed advanced data-mining algorithms in order to analyse their users' posts and activity. Such information was further used for two purposes: first, it was commercially sold to third parties; and second, it was fed into recommendation engines that then directed users to other relevant groups of interests. Finally, Experience Project also deployed emails and other notifications to notify users of activity in some of the groups to which they had belonged.

As such, the first and most essential question that the court faced was whether Ultimate Software was in fact eligible for immunity. The court ended up ruling that Ultimate Software was immune because its tools were content neutral: they facilitated communication but did not in and of themselves represent novel content. In reaching its decision, the court cited jurisprudence in Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1167–69 (9th Cir. 2008). In this case, Roommates.com asserted immunity against claims that it had violated the federal Fair Housing Act and California housing discrimination laws. In that case, it was ruled that Roomates.com did not qualify for immunity. Roomates.com, by requiring that its users provide information in the form of a limited set of pre-populated answers, became, in part, a developer of content or information.

Next, the court addressed the validity of the fourth claim: did Ultimate Software have a duty to warn its users of criminal activity that occurred on its platform? The plaintiff contended that Ultimate Software had a duty to inform her son that there was a criminal selling fentanyl-laced heroin on its platform. In the eyes of the plaintiff, this duty stemmed from the fact that Ultimate Software is like any brick-and-mortar business that has a special relationship with its clients and thereby an obligation to inform them of associated risks. More specifically, the plaintiff articulated that website operators are like

"restaurants, bars ... amusement parks and all businesses open to the public" and have the same duty that all businesses open to the public have to their invitees. Moreover, this duty, the plaintiff noted, involves taking affirmative action to control the wrongful acts of various third parties that could plausibly threaten invitees, in cases where the occupants have reasonable cause to anticipate such acts and the injury that might result from them. Ultimate Software in turn contended that websites have no duty to warn their users of the criminal activity of other users and that the plaintiff's son should have assumed the obvious risk of purchasing drugs from an anonymous Internet drug dealer.

However, the court took a further step and elucidated its own reasons why social network websites are not liable. First, the imposition of a duty would impose an ineffective general warning to all users. Second, it would have a chilling effect on the Internet by opening the possibility of further litigation. Third, the contention that brick-and-mortar businesses such as bars are similar to social networking websites is not persuasive. The allocation of risk partially depends on the degree to which harm is foreseeable, and the court contended that risk can be substantially more apparent in the real world than the virtual social network world. Moreover, even if Ultimate Software possessed superior knowledge of the sale of fentanyl-laced heroin, the possession of such knowledge only creates a special relationship if there is dependency or detrimental reliance by users. The court therefore rejected the plaintiff's argument that Ultimate Software should be liable because it owed a duty to her son.

This particular case was elected for inclusion as it concerns litigation around an issue that is only likely to become increasingly prominent: social media platforms using AI-informed tools to facilitate behaviour that can be potentially harmful for their users. This court's decision suggests that for a plethora of charges such social media platforms are eligible for immunity, unless they create and develop information or content that then facilitates harmful behaviour. Naturally, the question then becomes: what actually constitutes the creation of content? And in this case, the court ruled that mining insights from algorithms and then using those insights to steer users does not necessarily constitute information creation. Conversely, framing the creation of content in a way that, for instance, requires the submission of unlawful answers and steers users based on that information, as Roomates.com did in Fair Hous. Council of San Fernando Valley v. Roommates.com, constitutes content creation. There seems to be well-established jurisprudence that argues that providing neutral tools for navigating websites is fully protected by CDA immunity, unless the website creators use such tools for unlawful purposes. While jurisprudence in this domain seems to be well established, this question might plausibly become thornier in the future with the rise of generative AI that creates content such as text, photos and videos.

Moreover, the Court ruled fairly decisively that Ultimate Software had no duty to warn its users of potential danger because it owed no special obligation to that user. The court drew a distinction between brick-and-mortar businesses that, it argued, possessed such a duty. One of the motivating factors in this distinction was that such businesses could more accurately foresee risks than social media websites.

c. Cahoo v. Fast Enters, LLC (20 December 2020)

Cahoo v. Fast Enters is a case that involved five plaintiffs that began putative class action cases to recover damages resulting from the State of Michigan's Unemployment Insurance Agency (UIA). MiDAS was a rudimentary AI system that operated using logic trees and was created to identify discrepancies in various unemployment compensation records, automatically determine whether claimants had committed fraud and execute collection. The plaintiffs asserted that the system contains a plethora of failures: lack of human oversight, detection of fraud where none existed, provision of little or no notice to accused claimants, failure to allow administrative appeals and assessment of penalties against blameless individuals. Originally, the amended complaint listed twelve specific counts

against the companies and individuals that the plaintiffs believed had contributed to the state's development of this flawed system. However, the case has been whittled down through motion practice and interlocutory appeal. At this point, only one procedural due process claim remains.

The defendants in this case were FAST Enterprises LLC and CSG Government Solutions, two companies that assisted the UIA in the development of MiDAS. The defendants petitioned for a second round of motion to dismiss on the grounds that the plaintiffs could not establish Article III standing because: (1) there was a failure to highlight an injury in fact because their claims were not completely adjudicated by MiDAS; (2) the alleged injuries were not traceable to the defendants; and (3) several plaintiffs are precluded from bringing their claims because they did not disclose property interests in the cause of action during the respective bankruptcy proceedings. Therefore, the key issue in this case concerned whether the court had subject matter jurisdiction over the dispute.

Ultimately, the court contended that the plaintiffs did have adequate standing and that the defendants, by virtue of their assistance in the development of MiDAS, were legally liable. Although it was true that the plaintiffs' fraud cases were not entirely auto-adjudicated by MiDAS, the UIA still employed the flawed system to determine the existence of fraud. Moreover, in several cases, the system deemed that fraud was present when a plaintiff merely failed to properly respond to a question. In addition, the UIA did not provide sufficient notice to the plaintiffs of their fraud determination. Crucially, at least in this case, both defendants – FAST and CSG – worked closely with the UIA in developing MiDAS, so the injuries that the plaintiffs claim are actually traceable to them.

In order to fully understand this case, is it important to understand how MiDAS actually worked. When the MiDAS logic tree-informed system observed a discrepancy between employer-paid wages and employee-claimed eligibility, it automatically generated a questionnaire to the claimant. If the questionnaire was then not responded to in a timely manner, there was an automatic determination that the claimant knowingly misrepresented or concealed information. Once a fraud determination was made, MiDAS issued three separate notices. The first was a Notice of Determination that observed there being an issue at hand. The second was another Notice of Determination that notified the claimants that their actions misled or concealed information alongside an announcement that any active claims would be terminated. These determinations did indicate that the claimants had a right to appeal a determination of fraud in court within thirty days. However, these notices oftentimes did not notify claimants of their ability to file late appeals for good cause. Moreover, because there were frequent errors in notifying claimants, many claimants did not become aware of such an assessment until after the appeal deadlines passed. If the claimants did not appeal the determination within thirty days, MiDAS automatically sent a letter in which it demanded payment of restitution, penalty and an accrued interest. Ultimately, the UIA deactivated MiDAS after another court found that the system was error-prone.[66]

The court ruled that the defendants misrepresented the claims of the plaintiffs and that the occurrence of a sufficient injury in fact was demonstrated. First, the plaintiffs never contended that MiDAS deployed sophisticated AI tools. Rather, their contention was that fraud determinations were wrongfully made based on MiDAS's application of logic trees, which led to automated decisions that were unfair. Second, the core of the plaintiffs' complaints is that the automated system did not afford the kind of pre-deprivation process that state law required. Furthermore, the plaintiffs held that post-deprivation remedies were inadequate because that decision-making process lacked transparency. Therefore, even if it is true, as the defendants claim, that UIA employees participated in all of the fraud adjudications, the plaintiffs can still demonstrate an injury in fact given that they were deprived of unemployment benefits without adequate pre- or post-deprivation processes.

[66] Zynda v. Arwood, 175 F. Supp. 3d 791 (E.D. Mich. 2016).

This case is significant for many reasons. First, it is broadly indicative of the fact that companies and government agencies are often eager to deploy AI tools without first developing a complete understanding of how such tools will work and what their associated ethical problems may be. In this case, the UIA deployed a system that was error-prone, deployed the tool before these errors were fully understood and, as a result, facilitated the inequitable handling of several unemployment benefit claims. As AI grows in technical sophistication and more companies rush to develop AI-related tools, the possibility of those tools being misused along the lines of the UIA and MiDAS likewise increases. Both courts and legislative bodies will have important roles to play in working to ensure that AI tools do not perpetuate inequitable outcomes: courts can work towards ensuring that justice is served when inequitable outcomes result and legislative bodies can work towards imposing regulations to limit the occurrence of such outcomes.

This case also highlights what degree of involvement in the development of AI-related technologies is sufficient for establishing traceability in instances when injury has been committed. Advising is not sufficient for liability; however, more direct involvement in the creation and management of a tool is, even if the tool in question is deployed by another agency. This fact should serve as a warning to private entities that assist in the development of AI-related tools deployed by other actors, whether they are private or state: if the tool ultimately causes injury or violates some law, your company might be legally liable even though it did not directly oversee the tool's use.

Finally, this case interestingly sheds light on the types of adjudication processes that are deemed to be fair and legal in the case of unemployment insurance. The court ruled that it is essential that adequate pre- and post-deprivation processes are provided for. In this case specifically, it was shown that post-deprivation processes were inadequate because the decision-making process was not transparent. The AI system deployed in this case was admittedly rudimentary: MiDAS only made use of low-level logic trees. In the near future, it is possible to imagine systems that would be less prone to the types of errors that MiDAS committed and more able to flexibly consider the cases of unemployment insurance claimants. The possibility of their being superior systems gives rise to the following question: what degree of transparency is sufficient for adequate post-deprivation? Some neural network-based AI systems make decisions in a black-box fashion: with these systems it is sometimes not necessarily clear to the developers themselves how the systems reached the decisions they did. This particular legal case does not reveal what level of transparency would be sufficient on the part of an AI system in order to deem that it adjudicated an unemployment insurance claim transparently. As AI systems become increasingly advanced, this type of issue might be one for future courts to decide.

IV. Conclusion

In this paper, we provided insights into the reactions of judiciary and administrative institutions to the increasing deployment of AI systems in addition to a collation of the various responses to AI adoption and the guardrails developed to prevent abuse as a part of the overall regulatory efforts to reach a balance between achieving the benefits and minimising the risks associated with this powerful technology. The paper focused on select cases from the period between 2017 and 2022 in order to better capture the current status quo as reflected in a selection of rulings. Furthermore, the article briefly discussed emerging regulatory initiatives targeting AI and automated decision-making on both sides of the North Atlantic by way of presenting key facts and developments that can inform current and future debates with respect to regulatory frameworks applicable to AI systems.

The article focuses on select case law from the EU and the USA on subject areas that would likely be categorised as high-risk uses following the classification provided in the

draft EU AI Act. Some of the included cases would qualify as high risk as they concern the provision of essential private and public services; others are also deemed relevant in light of recent attention paid by the Members of the European Parliament to AI in the context of emotion recognition systems and the scraping of biometric data, both of which are items under discussion in the European Parliament.[67] The significant impacts that can result from government institutions deploying automated decision-making in sensitive areas, with a very large number of citizens potentially affected by biases or discriminatory algorithms, call for carefully designed rules and thorough assessments when relying on such systems in the public sector. Relevant examples are provided in the context of the SyRI case, which centres on algorithm bias, and the case of MiDAS in Michigan, where the tool's error-proneness was the central issue.

Although it is not possible to draw direct comparisons between the adjudicated cases across the Atlantic, our research centred on the analysis of actual judicial and administrative decisions regarding cases involving automated systems and aims to serve as a first attempt to inform the research and policy discourse on transatlantic governance efforts, taking into account relevant draft laws on AI and international cooperation efforts such as the EU–US AI roadmap. While regulatory efforts such as the European AI Act foresee extraterritorial application, the article discussed the decisions against Clearview AI under the GDPR, and it was noted that it can be quite difficult to operationalise extraterritoriality given enforcement challenges across jurisdictions. Hence, potential future mutually agreed-upon frameworks should also pay specific attention to operationalisation in addition to defining common values, such as explainability. Regarding the latter, achieving explainability to a satisfactory degree is not always straightforward, especially as AI systems become more complex and rely on deep neural networks and extensive unsupervised learning.

A growing number of economic actors are deploying AI technology in a variety of domains and incorporating this technology into an increasing number of practices. In the last few years, there has been a continuous decrease in cost per achieved computing performance. Greater affordability facilitates digital devices becoming increasingly prevalent in our daily lives and across our economy, which helps create an immense amount of data that can, in turn, serve as the basis for training or adapting algorithms. Cost decreases for computing performance also allow the development and training of increasingly potent AI systems. Together, these factors provide fertile ground for the dissemination and uptake of AI, with the accumulated impacts on economic activity and society as a whole being difficult to anticipate. Nevertheless, our case analysis indicates that algorithmic tools deployed by public authorities and by private companies in sensitive areas, where a significant number of citizens could potentially be affected by algorithm bias or procedural problems, require carefully designed rules and careful pre-deployment assessment. In light of these fast-paced developments, this article aims to offer a building block towards a better understanding of legal decisions regarding AI in the USA and in the EU to facilitate the work targeted at creating common frameworks.

Competing interests. The authors declare none.

[67] "AI Act: a step closer to the first rules on artificial intelligence" (2023) https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence (last accessed 20 May 2023).

Cite this article: E Kiesow Cortez and N Maslej (2023). "Adjudication of Artificial Intelligence and Automated Decision-Making Cases in Europe and the USA". *European Journal of Risk Regulation* **14**, 457–475. https://doi.org/10.1017/err.2023.61