Inventory of comparative judgments and rulings

AI and Vulnerable Groups

CONTENT

24 December 2024

1. Introduction

This study systematically analyses case law on the impact of AI on vulnerable groups. The research examines judicial and authoritative decisions globally, employing an intersectional approach to map vulnerabilities and propose safeguards against AI-related risks. Methodologically, it classifies cases by vulnerability type while addressing the challenges posed by the evolving legal landscape and limited precedent in this area.

1.1. Introduction to the study

Within the framework of Subgroup 1.2 of the Google Charity Project, our work focuses on tracking, compiling, and systematising the most relevant case law on the impact of AI on vulnerable groups. This includes identifying vulnerabilities and potential impacts on fundamental rights, with a particular focus on the Organisation for Economic Co-operation and Development (OECD) principles for the development and implementation of AI: Inclusive growth, sustainable development and well-being; Respect for the rule of law, human rights and democratic values, including fairness and privacy; Transparency and explainability; Robustness, security and safety; Accountability[1].

While the study focuses primarily on Europe and the Americas, it also includes relevant information from other regions of the world, such as Australia. The search prioritises judicial decisions, without excluding the analysis of other decisions by data protection authorities or similar bodies.

The following objectives are pursued:

- Provide information to quickly and concisely map the jurisprudence related to AI and vulnerable groups.
- Descriptive analytical study of vulnerable groups, the type of court or authority issuing decisions, the geographical area and the year in which the decision was issued.
- Assessment of judicial or authoritative approaches to the issue, evaluating the possibility of identifying a consistent legal standard regarding the risks of AI to vulnerable groups and the safeguards that must be in place when such systems are used, based on the decisions analysed.
- To formulate proposals for non-governmental organisations (NGOs) working with different vulnerable groups.

[1] OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449), Adopted on: 22 May 2019; Amended on: 03 May 2024, available: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

This report adopts a concept of vulnerability in the context of AI technologies that encompasses not only traditionally vulnerable groups, but also individuals who may be affected by AI at any given time. It recognises that AI systems may exploit human vulnerabilities, taking into account not only demographic and socio-economic conditions or psychosocial factors but also contextual, relational, situational, and temporal factors[2].

1.2. Methodology

The first task was to search for judgments using legal databases, doctrinal articles, reports from public and private organisations, news media and case tracking. The number of judgments obtained is not extensive because, despite the undeniable reality of the use of AI systems in many countries, legislation regulating their use and the legal consequences of their misuse is relatively recent. As a result, many incidents have not yet reached the courts, while others remain unresolved. Some of these incidents are discussed in the report because of their particular relevance and link to specific judgments analysed. In addition, in certain cases, decisions have been appealed, which will require attention for future decisions, as in the cases of BOSCO or WORKDAY.

The set of decisions has been classified on the basis of the following vulnerable groups: people in situations of poverty or social exclusion; informal and precarious workers; rural workers and rural populations; persons belonging to racial or ethnic minorities, including migrants, refugees and indigenous peoples; women and persons exposed to gender-based discrimination; religious, political or philosophical minorities; children and adolescents; older persons; people with physical, mental, sensory or intellectual disabilities; people with chronic illnesses or health conditions that lead to discrimination; people living with HIV/AIDS or other health conditions that make them vulnerable to discrimination; LGBTQ+ people and those whose gender expression does not conform to traditional expectations; and people who speak a minority language or do not speak the dominant language in their environment[3]. In the appendix you will find a summary table of each judgment, with the name of the case with its reference, the court or data protection authority from which it originates, the year in which the judgment was issued and the vulnerable group affected by the IA system.

It should be noted, however, that in many cases the group concerned is not categorised by a single factor, but by several. In this regard, an intersectional approach is particularly important as it highlights the complexity of addressing

[2] OTERO, B., *AI for Good: La idea de la vulnerabilidad humana en tela de juicio*, 2 october 2023, available at: https://www.odiseia.org/post/ai-for-good-la-idea-de-la-vulnerabilidad-humana-en-tela-de-juicio-1

[3] List for the cAIre project provided by OdiseIA.

these cases. This is illustrated by the cases of AFR Locate in the United Kingdom, as well as CRIMSAFE and WORKDAY in the United States, where multiple categories of vulnerability intersect.

2. Vulnerable groups and relevant judgements

Of all the vulnerable groups identified in the first section of this report, we have found judgments related to the use of algorithmic or AI systems in the following cases: people in situations of poverty or social exclusion; informal and precarious workers; persons belonging to racial or ethnic minorities, including migrants, refugees and indigenous peoples; women and persons exposed to gender-based discrimination; children and adolescents; older persons; people with physical, mental, sensory or intellectual disabilities; and people with chronic illnesses or health conditions that lead to discrimination.

The analysis of the identified judgments is structured as follows: the origin of the case, which outlines the facts, the decision made by the court or authority, and, finally, key findings regarding the role of AI in affecting vulnerable groups.

2.1 People in situations of poverty or social exclusion

In this subsection, we not only present the main judgments related to people in situations of poverty or social exclusion but also analyze three significant cases that, to the best of our knowledge, have not been brought before the courts.

2.1.1. ROBODEBT: automated-debt collection system, Australia

Date of final decision: June 11, 2021
Authority: Federal Court of Australia

- *Origin of the case*:

The applicants filed a class action on behalf of approximately 648,000 group members against the Commonwealth of Australia for its use of an automated debt-collection system. This system, colloquially known as the ROBODEBT system, was designed to recover overpaid social security payments from recipients.

Social security recipients who do not earn a constant fortnightly wage, do not earn a fortnightly income or only work for intermittent periods in a year were affected. Many people who were required to repay illegally declared debts could not afford to repay these amounts.

The applicants sought court approval of a proposed settlement of the class action.

- *Decision*:

The Commonwealth conceded, and the court found, that it did not have a proper legal basis to raise, demand or recover asserted debts based on income averaging from the Australian Taxation Office (ATO) data.

His Honour stated that the Commonwealth's failure was particularly acute given that many people who faced demands for repayment of unlawfully asserted debts could not afford to repay those amounts, insisting that recipients of social security benefits are particularly vulnerable and ill-equipped to properly understand or challenge the basis of the asserted debts.

The approved proposed settlement required that the Commonwealth pay $112 million, to be distributed proportionately amongst relevant group members depending on the size of their debt and how long they were without their money.

- *Key findings*

As the judge underlined, this proceeding has exposed a shameful chapter in the administration of the Commonwealth Social Security system and a massive failure of public administration.

The significance of the decision is that it puts governments on notice that they cannot merely rely on automatic systems and broad assumptions in formulating and implementing policy, particularly in the context of social welfare, and especially where the consequences are particularly felt by persons who are disadvantaged in terms of resources, capacity and information[4].

2.1.2. BOSCO: decision making tool, Spain

Date of final decision: April 30, 2024 (cassation appeal November 27, 2024)
Authority: Spanish National Court (Audiencia Nacional)

- *Origin of the case:*

BOSCO is an application developed by the Spanish government and provided to electricity companies. This tool is used to determine whether a vulnerable user qualifies for discounts on their electricity bill (verifying applicants' eligibility for the social electricity subsidy). Concerns have been raised about its accuracy, as the

[4] Humarights Lw Center, *The Federal Court approves a $112 million settlement for the failures of the Robodebt system*, 11 june 2021, available at: https://www.hrlc.org.au/human-rights-case-summaries/2021/9/30/the-federal-court-approves-a-112-million-settlement-for-the-failures-of-the-robodebt-system

tool has denied subsidies to individuals who were entitled to them, although this could not be proven due to a lack of transparency.

Due to its malfunctioning, the Civio Foundation requested access to the source code of the social electricity subsidy tool. While technical information was provided, access to the source code was denied by the Transparency Council (Resolution 701/2018, dated February 18, 2019), rejected again by Central Administrative Court Judgement 143/2021, dated December 30, and most recently by the National Court in a Judgement on April 30, 2024, which dismissed the administrative appeal. In other words, access has been denied on three separate occasions.

- *Decision:*

In the Judgement of April 30, 2024, the Civio Foundation was denied access to the source code of BOSCO for the third time. The National Court rejected Civio's arguments, stating that the source code is protected under Intellectual Property Law and that providing it would significantly endanger the rights of third parties and conflict with legally protected interests, as defined by the limitations on access to public information under Article 14 of Law 19/2013, dated December 9, on transparency.

- *Key findings*

This Judgement missed an opportunity to address the guarantees of due process and protections for individuals facing administrative procedures involving digital tools, rendering it weak from a doctrinal and jurisprudential standpoint. The case underscores the importance of transparency in the use of algorithms by public administrations and the need for regulations that ensure such transparency. Updating transparency legislation is essential, particularly to impose and define proactive disclosure in the form of public algorithm registries[5].

The Civio Foundation has filed a document to prepare for an appeal before the Supreme Court.

The cassation appeal was granted by the Administrative Chamber of the Supreme Court (First Section) on 27 November 2024. The Court stated that the issue raised in the appeal was of objective cassation interest for the development of case law, in particular for determining whether it is appropriate to provide the source code of the software application in order to verify compliance with the requirements to qualify for the social bonus.

[5] COTINO HUESO, L., "Caso Bosco, a la tercera tampoco va la vencida. Mal camino en el acceso a los algoritmos públicos", *Diario LA LEY*, nº 84, Sección Ciberderecho, 17 May 2024, p. 5.

The case has been referred to the Third Section of the Administrative Chamber of the Supreme Court for consideration of the appeal. We are currently awaiting its final decision.

2.1.3. CalWIN: automated decision system, California

Date of final decision: November 13, 2013
Authority: Court of Appeal of the State of California, Third Appellate District (Sacramento)

- *Origin of the case:*

The case arose when the plaintiffs, who were welfare recipients in California, claimed that the CalWIN system automatically and erroneously terminated or delayed their benefits without sufficient human oversight or intervention. This led to a lawsuit in which the plaintiffs sought a writ of mandate to compel the California Department of Social Services to modify the system to prevent these automatic actions. They claimed that the system's failure to correctly process required eligibility reports resulted in unwarranted terminations, reductions or delays in welfare benefits.

- *Decision:*

The court ruled in favour of the Department of Social Services, confirming that the termination or reduction of benefits was the result of human error rather than systemic flaws in the CalWIN system itself. The court also affirmed that the department maintained sufficient oversight of the system and exercised proper discretion in administering it. The trial court's decision to sustain a demurrer and dismiss the claims was affirmed without leave to amend.

- *Key findings*

The CalWIN system embodies a form of automation and algorithmic decision-making that affects vulnerable groups, particularly low-income individuals and families who rely on welfare programmes. The court acknowledged the potential for systemic error, but attributed the primary problem to caseworker error rather than inherent flaws in the automation process. The judgment pointed out the importance of proper oversight, human intervention and training to ensure that vulnerable recipients are not unfairly denied benefits due to errors in automated systems.

2.1.4. OFQUAL: automated decision system, United Kingdom

Date of final decision: August 5, 2021
Authority: Information Commissioner's Office (ICO)

- *Origin of the case:*

A-level and General Certificate of Secondary Education (GSCE) examinations couldn’t take place because of Covid-19. Hence, OFQUAL (the regulator of qualifications, exams and tests) tasked an algorithm with assigning grades.

The complainant wrote to OFQUAL and requested statistical information relating to the adjustments made to A-level grades, based on what is known as ‘the algorithm’, and contacted the Commissioner (ICO) to complain about the way that their request for information had been handled.

There were concerns that the algorithm itself was unlawful, not only breaching anti-discrimination standards but also Article 22 of the GDPR which outlines the right not to be subject to fully automated decision making that significantly affects individuals. The complainant has made this request based on concerns that students attending lower- performing centres from more deprived areas were disadvantaged by the algorithm.

- *Decision:*

The Commissioner decided that the public interest lies in disclosure. Disclosure would seek to build a bigger picture of a process that delivered “significant inconsistencies” and will demonstrate how justified and widespread concerns regarding the algorithm were. The Commissioner concurs with OFQUAL when it says that disclosure would be likely to have repercussions. However, the Commissioner disagrees with the prejudice to current students that OFQUAL has foreseen and, with this in mind, orders disclosure.

OFQUAL appealed to the First-tier Tribunal. The tribunal dismissed the appeal, but OFQUAL had permission to appeal to the Upper Tribunal (Administrative Appeals Chamber). The case is now back in the hands of the First-tier Tribunal to clarify the public interest issue.

- *Key findings:*

Machine learning made a prediction based on historical data, ending up reinforcing existing inequalities. The complainant's request for information was based on concerns that students from centers in more economically disadvantaged areas had been harmed by the algorithm, and disclosure of the information would serve to hold OFQUAL accountable.

This reflects a broader concern about the fairness of the process of assigning grades based on historical school performance. These inequities in the grading outcome fueled public debate about the fairness of OFQUAL’s algorithm and decision as evidenced by the resolution: “OFQUAL recognises that disclosure would illustrate the variances of adjustments that the algorithm made to Certificate of Advanced Graduate Study (CAGs) across centres in England […] the algorithm saw almost 40% of students in England, Wales and Northern Ireland awarded a grade lower than their CAG and was met with widespread criticism within the mainstream media […] The complainant has made this request based on concerns that students attending lower-performing centres from more deprived areas were disadvantaged by the algorithm”.

2.1.5. SCHUFA: credit scoring and automated decision making, Germany

Date of final decision: December 7, 2023
Authority: European Court of Justice

- *Origin of the case:*

SCHUFA is a company that provides creditworthiness information to its clients, including banks and lenders, through credit scoring processes. Credit scoring is the process of assigning a score to a credit applicant based on mathematical and statistical models. This score is generated from credit profiles created using data from individuals with similar characteristics.

The plaintiff in this case is an individual whose credit application was rejected based on the information provided by SCHUFA, which was subsequently passed on to the lending institution. The plaintiff exercised his right of access to personal data and requested detailed information from SCHUFA. However, SCHUFA only provided general information, citing trade secrets relating to the profiling algorithm. It refused to disclose the individual’s specific data or the weighting used to generate the score. SCHUFA argued that the final decision to grant or deny credit rested with its contractual partners (the lenders) and that SCHUFA merely provided the credit score.

The Wiesbaden Court (Germany) referred a preliminary question to the European Court of Justice (CJEU) on the interpretation of Articles 6 and 22.1 of the General Data Protection Regulation (GDPR). The question focused on whether the automated generation of credit scores by a credit reference agency (such as SCHUFA) falls within the scope of Article 22 of the GDPR, given that SCHUFA itself does not make the final automated decision, but merely provides the credit score to its partners.

- *Decision:*

In its judgment of 7 December 2023, the CJEU analysed for the first time Article 22 of the GDPR on automated decision-making. The Court concluded that the automated generation of a probability score by a credit reference agency, based on personal data about an individual's ability to meet future financial obligations, constitutes an "automated individual decision" under Article 22(1). This is the case where the score plays a determining role in a third party's decision to enter into, perform or terminate a contract with the individual.

The Court interpreted Article 22 broadly, holding that a score generated from a probability value is a “fully automated decision” if it significantly influences the decision of a third party, such as a lender, in entering into a contractual agreement.

- *Key findings:*

The judgment extends the scope of Article 22 of the GDPR beyond the formal decision-maker, to include third parties that process data. Furthermore, the judgment highlights the significant impact that automated decisions can have on the outcome, even if the decision appears to have been made by a human or other entity. This approach represents a significant advance in legal protection against the risks posed by automation and AI[6].

2.1.6. SyRI: fraud risk assessment system, the Netherlands

Date of final decision: February 5, 2020
Authority: The Hague District Court (Rechtbank Den Haag)

- *Origin of the case:*

According to the Court, the Systeem Risicoindicatie (better known as SyRI) is a legal instrument used by the Dutch government to prevent and combat fraud in areas such as social security, income-dependent schemes, taxes, social security, and labour laws.

The system processes an almost unlimited amount of structured data from existing and available records. A total of 17 categories of data are eligible for use, including employment, administrative sanctions, tax information, movable and immovable assets, social benefits, business data, housing, identification, civic integration, compliance with legislation, education, pensions, debts, permits, and exemptions, as well as information on whether a person is covered by the Health

[6] COTINO HUESO, L., “La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial”, *Diario LA LEY*, nº 80, Sección Ciberderecho, 17 January 2024.

Insurance Act (Section 4.17)[7]. SyRI feeds this mass of data into a risk model using fixed indicators, which then generates a list of citizens deemed to have a higher risk of fraud. The data may only be used to produce a risk report on a natural or legal person who is considered worthy of investigation for potential fraud, unlawful use, or non-compliance with regulations. The Social Affairs and Employment Inspectorate then analyzes the risk report produced by SyRI.

- *Decision:*

The court determined that this legislation fails to comply with Article 8 of the European Convention on Human Rights (ECHR), which safeguards the right to privacy in personal and family matters, as well as home and correspondence. The judgment stated that the legislation does not achieve a fair balance, as required by the ECHR, to sufficiently justify any violation of private life, nor does it adequately prevent the risk of discrimination. The implementation of SyRI was found to lack sufficient transparency and accountability. As a result, the court declared the SyRI legislation unlawful and non-binding, as it contravenes higher legal standards.

- *Key findings:*

This case is not only significant for the protection it affords to the fundamental right to privacy (right to data protection) under Article 8.2 of the ECHR but also for the transparency issues it exposed. The court noted that “it is hard to imagine any type of personal data that is not eligible for processing in SyRI” (Section 6.98).

A clear lack of information was identified regarding how the risk model works, the types of algorithms used, and the risk analysis methods applied in the second phase by the Inspectorate[8]. As stated in the judgment, this lack of transparency could lead to discriminatory outcomes based on biases such as lower socio-economic status or an immigrant background. The Dutch government itself admitted that SyRI had only been used in so-called “problem districts” (Section 6.93), and the UN Special Rapporteur on extreme poverty and human rights has warned about the discriminatory and stigmatizing effects of SyRI (Section 6.92).

The following outlines similar cases to SyRI that have not reached judicial instances.

[7] COTINO HUESO, L., “Holanda: SyRI, ¿a quién sancionó? Garantías frente al uso de la inteligencia artificial y decisión automatizada en el sector público y la sentencia holandesa de febrero de 2020 (1)”, *La Ley Digital* 4999/2020, de 18 February de 2021, p. 8.

[8] OUBIÑA BARBOLLA, S., “Límites a la utilización de algoritmos en el sector público: reflexiones a propósito del caso SyRI”, *Justicia algorítmica y neuroderecho: una mirada multidisciplinary* (BARONA VILAR, ed.), Tirant Lo Blanch, Valencia, 2021. p. 659.

- PAMAS: Job profiling system, Austria

In 2020, following a pilot phase, the Austrian Public Employment Service (*Arbeitsmarktservice,* AMS) implemented a statistical profiling system called PAMAS (*Personalised Arbeitsmarktchancen Assistenzsystem*) to assess jobseekers' prospects in the market. This algorithm evaluates various characteristics of unemployed individuals and assigns each person a score. Based on this statistical model, individuals are classified into three categories - high, medium, or low likelihood of finding a new job- each receiving different levels of support for labour market reintegration.

Controversy surrounding PAMAS emerged when the model's source code, published by the contracted development company, revealed point deductions for certain disadvantaged or vulnerable groups. For instance, points were subtracted for being a woman or a non-EU citizen. Consequently, individuals belonging to multiple vulnerable groups, especially those at higher risk of social exclusion, faced significant score reductions, often placing them in the "low likelihood of finding a job" category. This raised concerns about potential bias and discriminatory impacts embedded within the automated system.

The discrimination does not stem from the algorithm itself but from the human decisions made based on its results. The system accurately identified that factors such as being a woman or a person of color were statistically associated with a lower probability of finding employment. The issue lies in the programming choices, where the system was designed to allocate fewer resources to individuals with lower-probability outcomes, effectively reinforcing patterns of social exclusion. This human decision to prioritize resources according to these statistical outcomes is what ultimately produced the discriminatory impact.

While this system is objective in that it mirrors existing discriminatory practices in the labour market, it has faced significant criticism from various Austrian organizations and social sectors. These critiques are well-founded, as the system contributes to the stigmatization of vulnerable groups by categorizing their members in the lowest employment probability bracket. This classification reinforces negative stereotypes and further marginalizes those already at a disadvantage, perpetuating barriers to their reintegration into the workforce[9].

The public unemployment service defended this classification approach by claiming it would allow for better support to individuals facing greater challenges

[9] SORIANO ARNANZ, A., "Decisiones automatizadas y discriminación: aproximación y propuestas generales", *Revista General de Derecho Administrativo*, nº 56, enero 2021.

in finding employment. However, it ultimately prioritized efficiency in the allocation of public resources over other considerations. Concluding that the most efficient use of resources would be to channel more support towards individuals with an average probability of finding a new job, the service significantly reduced the resources and assistance allocated to those with a lower likelihood of re-entering the labour market. This decision, in turn, perpetuated social exclusion for certain vulnerable groups who received less support, reinforcing the very barriers they already faced[10].

- Automated scoring system profiling in labor offices, Poland

In 2014 the Polish government introduced a profiling mechanism for unemployed individuals that was supposed to allow support to be tailored to individual needs and reduce bureaucratic inefficiency.

The system worked as follows: unemployed individuals were classified into three groups according to their proximity to securing employment. Each group received a specific type of assistance tailored to their situation. This categorization was intended to be semi-automated, utilizing a scoring system that assigns each person to one of the three profiles based on 24 data factors.

Based on the final score the algorithm decided which category shall be given to the unemployed person. This determined the scope of assistance that a person can apply for. The third category (containing around 30% of the unemployed, those facing serious difficulties like chronic disease, disability or addiction), in theory, were supposed to be able to apply for some sort of assistance. In practice, financial and organizational problems mean local job centres offer little to those in this category. They end up being written off as a helpless group that is not worth investing in[11].

The profiling mechanism was originally intended as a guidance tool, allowing staff to have the final decision on which group an individual should be placed in. Once a person’s profile is calculated, the system enables clerks to either accept or reject the computer's decision. However, early statistics suggest that clerks chose to override the result in fewer than 1 out of every 100 cases.

After numerous criticisms, in 2019 the government decided to end its experiment with profiling the unemployed. The primary concerns have centered around the lack of transparency in how decisions are made. Lack of transparency in the

[10] Ibid.

[11] NIKLAS, J., “Poland: Government to scrap controversial unemployment scoring system”, Algorithm Watch, 16 April 2019, available at: https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/

process of profiling is directly related to the choice of the computer system as the main decision-making tool and the decision to keep the underlying algorithm secret (even from the frontline staff who are responsible for carrying out the interview with the unemployed) [12]. Unemployed individuals had no right to access information on how the computer system determines their status, including the logic behind it, the specific data used, or how it impacts the final decision. Additionally, they were unable to challenge the computer's decision or request human intervention in the process.

The second concern relates to the risk of discrimination. Assignment to a specific profile is based on factors such as age, gender, and disability status, meaning that the situation of certain unemployed individuals, particularly their actual access to labor market programs, is influenced by these characteristics.
The Supreme Audit Office (Najwyższa Izba Kontroli), responsible for overseeing the state budget, public spending, and the management of public assets, conducted a comprehensive review of the profiling mechanism. The review revealed that the system is ineffective and potentially discriminatory. Under the scoring criteria, women are evaluated differently than men, and individuals from vulnerable groups, such as single mothers, people with disabilities, and rural residents, are more likely to be placed in the third profile.

- Toeslagenaffaire, the Netherlands

The Dutch childcare benefits scandal, or Toeslagenaffaire, illustrates the adverse impact of unregulated artificial intelligence on vulnerable populations. Beginning in 2013, the Dutch tax authorities implemented an algorithmic risk classification model to identify and combat childcare benefits fraud. This system, employing self-learning algorithms, flagged applications for fraud investigations based on a variety of criteria, including nationality, leading to systemic racial profiling[13].

The algorithm labeled non-Dutch nationals and individuals with dual citizenship as high-risk, subjecting them disproportionately to audits and severe financial penalties. Many families, often from immigrant or minority backgrounds, were falsely accused of fraud for minor administrative errors or omissions. These errors resulted in the suspension of benefits and demands for immediate repayment of

[12] FUNDACJA PANOPTIKON (Jędrzej Niklas, Karolina Sztandar-Sztanderska, Katarzyna Szymielewicz), *Profiling the unemployed in poland: social and political implications of algorithmic decision making*, Warsaw, 2015, available at: https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf

[13] AMNESTY INTERNATIONAL, X*enophobic Machines - Discrimination through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal*, London, available at: https://www.amnesty.org/en/documents/eur35/4686/2021/en/

large sums, pushing families into debt, unemployment, and homelessness, with severe psychological and social consequences.

The system's opacity and lack of accountability compounded the problem, making it impossible for affected families to challenge the decisions effectively. Despite prior warnings about the human rights risks of such automated systems, no measures were taken to prevent discrimination or ensure transparency. The revelations of these practices led to a national scandal, the resignation of the Dutch government in 2021, and calls for compensation and reform.

2.1.7. DUB & BRADSTREET Austria: automated decision-making, Austria

Date of final decision: September 12, 2024
Authority: Opinion of Advocate General of the Court of Justice of the European Union (CJEU), Mr. Richard de la Tour

- *Origin of the case:*

The case arose when a mobile operator refused to conclude or renew a mobile phone with an individual (CK), contracting insufficient financial solvency. This assessment was based on a credit evaluation performed by DUB & BRADSTREET Austria GmbH (formerly Bisnode Austria GmbH). CK, seeking to understand the reasons behind the refusal, submitted a claim to the Austrian Data Protection Authority requesting information about the logic applied in the automated decision-making process. The Authority granted CK’s request, but due to subsequent appeals, the Vienna Regional Administrative Court referred the matter to the Court of Justice of the European Union (CJEU). The key issue was to determine the extent of the information that DUB & BRADSTREET must disclose to CK to ensure transparency and accuracy in the automated credit assessment process.

- *Opinion of the Advocate General of the CJEU:*

The Advocate General of the CJEU opined that, under Article 15(1)(h) of the GDPR, DUB & BRADSTREET is obligated to provide CK with "meaningful" information about the logic of the automated decision-making process. This includes the methodology and criteria used to evaluate CK’s creditworthiness but excludes complex technical details, such as complete algorithms, when they qualify as trade secrets or involve the personal data of third parties.

The Advocate General emphasized that the information provided must be concise, accessible, and easily understood. The goal is to allow CK to verify the accuracy of the decision and understand the causal connection between the methods used and the outcome. This ensures that the data subject can fully comprehend how their creditworthiness was assessed and identify any potential inaccuracies.

- *Key findings:*

This case underlines the tension between transparency in automated decision-making and the protection of intellectual property and third-party data. The Advocate General's interpretation ensures that the right to meaningful information under the GDPR is not diluted by overly technical or opaque explanations.

The findings underscore the need to protect vulnerable groups from opaque algorithmic decisions that can significantly impact their fundamental rights. The Advocate General affirmed that balancing transparency with the protection of intellectual property must not be used as a pretext to deny individuals their right to understand decisions affecting them. Courts or competent bodies are tasked with weighing these rights and ensuring fairness.

The case sets a precedent for ensuring transparency in automated decision-making, providing a safeguard for individuals, including those from vulnerable groups, to challenge and verify the fairness of algorithmic outcomes.

2.1. Informal and precarious workers

This section analyses three cases of informal and precarious workers and a fourth case of workers who, although not precarious, are negatively affected by an algorithm that decides the location of their jobs, with the adverse consequences that this entails.

2.2.1. AMAZON: Flex delivery app, United States

Date of final decision: Pending resolution
Authority: Cobb County State Court, Georgia, USA

- *Origin of the case:*

The accident occurred on 15 March 2021 in Marietta, Georgia, when Rana, a passenger in a Tesla vehicle, was rear-ended by an AMAZON driver, Bryan Williams, who allegedly failed to exercise due care and was reportedly under the influence of drugs. AMAZON and Harper Logistics, acting as AMAZON's Delivery Service Provider (DSP), are identified as liable parties by virtue of their vicarious employment relationship with Williams and the implementation of an algorithmic system that pressures drivers to meet unrealistic objectives.The lawsuit was filed on October 22, 2021.

- *Decision:*

The case remains pending and awaits judicial resolution. The claim asserts multiple causes of action, including negligence, negligence *per se*, and vicarious liability under the principle of *respondedat superior*. The complaint argues that AMAZON, through its Flex delivery algorithm and other AI systems, imposes delivery targets that incite drivers to act recklessly. Furthermore, AMAZON exercises extensive control over its DSPs and drivers, challenging the DSPs' corporate independence and raising the potential for corporate veil piercing to hold Amazon directly accountable.

- *Key findings:*

AMAZON's reliance on artificial intelligence and supervisory algorithms lies at the heart of this legal claim that sheds light on the profound implications of technology on corporate liability and public safety. Central to the case is the company's Flex system, an AI-driven application that governs critical aspects of delivery operations, including route allocation, timing, and real-time monitoring of drivers' speed and location. While the system is designed to optimize efficiency, the claim argues that it imposes such stringent performance targets that it undermines the safety of both drivers and the public. By prioritizing speed over caution, Amazon's approach allegedly creates a hazardous work environment.

This pressure is most evident in the "rabbit speed" threshold, a metric used to evaluate delivery speed and efficiency. Drivers are compelled to meet this AI-monitored benchmark or face penalties, such as damage to their FICO score, which directly affects their earnings. According to the claim, this relentless algorithmic pressure incentivizes drivers to adopt unsafe practices, including driving at dangerously high speeds. Such behavior, it is argued, directly links Amazon's operational model to potential liability for compromising safety standards.

Adding to the complexity is AMAZON's dual role in training and supervising its drivers. While drivers are technically employed by third-party delivery service partners (DSPs), AMAZON provides initial training and establishes supervisory guidelines. In this case, Williams, the driver involved, was employed by Harper Logistics, but AMAZON conducted his background check and approved his employment. This involvement complicates the question of accountability, suggesting that AMAZON's oversight (or lack thereof) might render it partially liable for inadequate supervision.

The legal claim seeks substantial compensation for the devastating consequences of these alleged shortcomings. Rana, the injured party, has endured severe physical and emotional harm, leading to significant medical expenses, pain and suffering, and lost earnings. In addition to these damages, the claim seeks punitive compensation, citing Amazon's aggravated negligence in ignoring the risks posed by its algorithmic systems to both public safety and the welfare of its drivers.

This case serves as a critical example of the intersection between technology and corporate liability, particularly in contractor relationships. The claim points out AMAZON's near-total control over its DSPs and drivers through AI systems, potentially setting a transformative precedent. As the legal system grapples with these emerging dynamics, the outcome could redefine the boundaries of corporate responsibility and employment relationships in the context of AI-driven operations.

2.2.2. UBER: employment in the digital era, United Kingdom

Date of Final Decision: December 19, 2018
Authority: Court of Appeal (Civil Division), United Kingdom

- *Origin of the case:*

The case originated from a claim filed by UBER drivers in London, seeking recognition as "workers" under UK employment legislation. This recognition would afford them rights under the Employment Rights Act 1996 and the National Minimum Wage Act 1998, including entitlements such as minimum wage and compensation for working hours. The claimants argued that, despite their designation as "independent contractors", they were effectively subject to Uber's control through its platform and algorithms, resulting in a relationship of dependency.

- *Decision:*

The Court of Appeal upheld the decision of the Employment Tribunal, which had determined that the drivers qualified as "workers" by virtue of their relationship with UBER London Ltd. It found that UBER's control, primarily exercised through the app and its algorithms, limited drivers' autonomy in areas such as accepting and rejecting trips, performance ratings, and the use of the app itself. The decision establishes that, although the contracts designated the drivers as independent contractors, UBER's system structure imposed conditions that rendered them dependent on and subordinate to the company's decisions, which is incompatible with the autonomy expected in an independent contractor relationship.

- *Key findings:*

The pervasive influence of AI and algorithmic tools within the context of contemporary labour relations has come under judicial scrutiny, particularly in cases involving technology-driven platforms such as UBER. In the present matter, the algorithms integrated into UBER's application were identified as pivotal instruments of managerial control over drivers. These mechanisms, extending from the assignment of trips to the continuous monitoring and assessment of performance, were deemed to constitute a form of algorithmic supervision. The tribunal concluded that such oversight exemplified a hierarchical and subordinate relationship, an essential criterion for defining the employment status of workers.

Moreover, the tribunal recognized the economic and operational dependency of the drivers on UBER. It was established that the platform, through its algorithmic systems, exercised extensive control over access to customers, fare determination, and the operational conditions under which drivers performed their duties. This structure significantly restricted the drivers' capacity to operate as independent agents, reinforcing their reliance on the platform and underscoring their lack of autonomy.

The decision further elucidated the interpretation of the term “worker” under the framework of UK labour law. It affirmed that the classification of an individual's employment status transcends the language used in written contractual agreements. Instead, it is grounded in the practical realities of the relationship between the parties involved. Despite the contractual designation of UBER drivers as independent contractors, the tribunal found that the substantive dynamics of the relationship were marked by subordination and economic dependency, hallmarks of a worker classification.

This case also highlights a broader imperative to modernize labour protections in response to the rise of digital platforms and artificial intelligence. As these technological innovations reshape the contours of employment relationships, they pose significant challenges to the preservation of workers' rights. The tribunal's findings underscore the necessity of safeguarding these rights against contractual frameworks that may misrepresent or obscure the true nature of the employment relationship, ensuring that labour laws remain robust and equitable in the digital economy.

2.2.3. DELIVEROO: algorithmic discrimination and labour rights, Italy

Date of Final Decision: November 27, 2020
Authority: Ordinary Court, Bologna, Italy

- *Origin of the case:*

The conflict arose from the conditions governing riders' access to work sessions, which were managed through DELIVEROO's digital platform. The claim, lodged by the trade unions Filt Cgil, Filcams Cgil, and Nidil Cgil, alleged that the algorithm managing access to work sessions created discriminatory treatment against workers who participated in union actions, exercised their right to strike, or were unable to attend shifts for legitimate reasons such as illness or family responsibilities. The algorithm penalized these workers by lowering their scores, which negatively affected their ability to secure future shifts, creating a barrier to employment.

- *Decision:*

In its analysis, the Court of Bologna emphasized that DELIVEROO's system not only organized work through the digital platform but also created disparities in access to work sessions based on a reliability ranking. This ranking, determined by adherence to pre-booked sessions and participation during peak hours, placed riders in a position of dependence on the algorithm. The court recognized that this system created a significant disadvantage for riders who were compelled to cancel work sessions due to circumstances beyond their control, such as participating in strikes or personal situations justifying their absence.

DELIVEROO argued that the booking system was optional and that riders were not obliged to accept shifts. However, the court found that, in practice, the system severely impacted workers who did not meet the algorithm's requirements, as their ability to book future shifts was drastically reduced. The platform failed to consider legitimate reasons for cancellations, leading to discriminatory behaviour towards certain workers, particularly those exercising their right to strike.

- *Key findings:*

The court, in its deliberations on the employment practices of DELIVEROO, identified the algorithm employed by the platform as a mechanism of indirect discrimination. This system, designed to prioritize certain riders based on their ability to fulfil pre-booked work sessions, disproportionately disadvantaged individuals in vulnerable circumstances, including those with family responsibilities or health issues. Notably, the algorithm also imposed penalties on riders who engaged in strike actions, thereby encroaching upon the constitutionally enshrined right to strike, a cornerstone of labour rights.

The impact of such algorithmic technologies extends beyond mere organizational efficiency, reaching into the core of labour rights protections. The court acknowledged that while the algorithm ostensibly functioned in a neutral manner, it perpetuated disparities in access to work by failing to account for legitimate reasons behind certain riders' inability to participate in scheduled sessions. This case illustrates the latent capacity of algorithmic systems to engender discriminatory outcomes, even absent explicit intent, thus raising significant concerns about their broader implications in employment contexts.

In its findings, the court affirmed the responsibility of digital platforms to uphold labour rights and align their operations with anti-discrimination legal frameworks. DELIVEROO's reliance on the presumed neutrality of its algorithm was deemed insufficient to absolve it of accountability for the adverse effects such systems had on its workers. The judgment highlights that digital platforms must take proactive measures to mitigate the discriminatory potential of their technologies and ensure compliance with established labour standards.

Furthermore, the court reaffirmed the necessity of safeguarding union rights, particularly the right to strike, against the encroachments of algorithmic oversight. DELIVEROO's failure to adapt its algorithm to accommodate legitimate absences, such as those stemming from strike participation, constituted a violation of workers' rights. This discriminatory treatment, rooted in the rigid application of algorithmic logic, underscores the imperative for digital platforms to recognize and address the human realities underlying their operational systems, thereby fostering an equitable and rights-compliant employment environment.

2.2.4. TEACHER ALLOCATION ALGORITHM, Italy

Date of final decision: April 8, 2019
Authority: The Council of State in the Courts (Sixth Section) [Il Consiglio di Stato in sede giurisdizionale (Sezione Sesta)]

- *Origin of the case*:

In Italy, the Council of State dealt with a dispute concerning the use of an algorithm in the allocation of posts for teachers who participated in a national recruitment process in 2015. The applicants, after indicating their preferences on educational level and geographical location, noted that these were not respected. In contrast, other applicants did obtain places according to their preferences, questioning the fairness of the process and suggesting the absence of a meritocratic criterion in the allocations. The lack of transparency about the functioning of the algorithm and the lack of justification for the decisions led the appellants to question the legality of the system and to request that the algorithm be made public in order to assess its compliance with administrative law.

- *Decision*:

The judgment found that the algorithms violated (i) the principle of publicity and transparency due to the absence of information on the functioning of the algorithm without the affected and interested parties being aware of the criteria used and the data processed; (ii) the principle of merit and capacity infringing the principle of transparency because the allocation of places was carried out without providing the necessary information without respecting merit as some applicants were placed in unrelated places on the basis of their preferences or abilities, while others with lower scores were given preferential destinations; (iii) the principle of administrative motivation because the lack of justification in the automated decisions infringed the applicants' right to understand the reasons for the decisions and deprived them of a sufficient basis for appeal.

The court indicated that algorithmic decisions should be cognizable and reviewable, that the administration should verify the legality of the processes, allow judicial control at all stages, and ensure their comprehensibility. Furthermore, the Court ordered the Ministry of Education to re-evaluate the allocations on the basis of the preferences indicated in the teachers' rankings.

- *Key findings:*

This case is an example of how an AI system used in administrative decision-making, without adequate guarantees such as transparency or control and review mechanisms, can particularly affect vulnerable groups or people with less capacity for response or resources. In this instance, the system left teachers without any means to challenge arbitrary decisions.Lack of access to the logic behind the algorithm and the absence of clear and transparent criteria can lead to inequalities, especially for those who depend on the administration for their employment. It underlines the importance of ensuring clear and accessible accountability mechanisms and access to information to avoid or mitigate disproportionate impacts. Moreover, when automated decisions are not reviewable and understandable, those affected may have their right to defence undermined or eliminated, a situation that reflects the need for regulation to ensure transparency and fairness in AI-based technology.

2.3. Persons belonging to racial or ethnic minorities, including migrants, refugees. Indigenous peoples.

2.3.1. AFR Locate: automated facial recognition, United Kingdom

Date of final decision: August 11, 2020

Authority: Court of Appeal (Civil Division) on Appeal from the High Court of Justice Queen's Bench Division (Administrative Court)

- *Origin of the case:*

The case originated with Edward Bridges, a civil liberties campaigner, who challenged the use of Automated Facial Recognition (AFR) technology by South Wales Police (SWP) in an ongoing trial of a system called AFR Locate. AFR Locate involves the deployment of surveillance cameras to capture digital images of members of the public, which are then processed and compared with digital images of people on a watch list compiled by the SWP for the purpose of the operation. In the facts of this case, AFR Locate was used in an overt manner, not as a form of covert surveillance.

Bridges argued that the use of this technology breached his rights under Article 8 of the European Convention on Human Rights (right to privacy), data protection laws and the Public Sector Equality Duty (PSED) under the Equality Act 2010.

- *Decision:*

The Court of Appeal ruled in favour of Edward Bridges, stating that SWP's use of AFR technology was unlawful for the following three reasons: i) Lack of an adequate regulatory framework: The Court concluded that the legal framework governing the use of AFR was not adequate to ensure that the technology was used in accordance with the law, as required by Article 8 of the ECHR; ii) Data protection issues: The Court found that the use of AFR constituted "sensitive processing" under the Data Protection Act 2018, and that SWP had failed to meet the strict requirements for such processing; ii) Public Sector Equality Duty: The Court found that SWP had failed to adequately consider the potential bias of the AFR technology, in particular its impact on gender and ethnic groups.

The Court stressed that all users of AI systems must be fully aware of the mechanism and operation of the machine, as well as the data on which it bases its decisions. Lack of knowledge or partial knowledge of the AI system on the part of the user is sufficient to establish liability.

- *Key findings:*

The case highlighted concerns about potential bias in AFR technology, particularly in generating higher false positive rates for women and ethnic minorities. It highlighted the risk that AI could have a disproportionate impact on vulnerable groups, raising questions about indirect discrimination and the adequacy of the safeguards in place to protect these communities.

The Court notes that "SWP have never sought to satisfy themselves, either directly or by way of independent verification, that the software program in this case does not have an unacceptable bias on grounds of race or sex". It also points out, in the context of commercial confidentiality, the importance of public authorities having access to relevant information about the operation of the system in order to avoid indirect discrimination on grounds of race or sex.

2.3.2. COMPAS: risk assessment tools for recidivism, United States

Date of final decision: July 13, 2016
Authority: Wisconsin Supreme Court

- *Origin of the case:*

The COMPAS system (Correctional Offender Management Profiling for Alternative Sanctions) is a tool designed to assess recidivism risk and is used in several U.S. states. In this notable case, Mr. Loomis was sentenced to six years in prison. The court based its decision, in part, on COMPAS's assessment, which indicated a high likelihood of reoffending, ultimately denying him probation.

Mr. Loomis appealed the judgment to the Wisconsin Supreme Court, arguing that his right to a fair trial had been violated. He claimed that the defense was unable to challenge the methods behind the COMPAS software, as its algorithm was a proprietary secret. Additionally, Loomis contended that the system resulted in a non-individualized sentence and that its methodology introduced gender bias.

- *Decision:*

In 2016, the Wisconsin Supreme Court upheld the use of predictive algorithms like COMPAS in recidivism assessments, affirming that its application did not breach due process. However, the court stressed that such tools should not serve as the sole basis for sentencing. The court outlined that while COMPAS cannot dictate the severity of a sentence, it can be a relevant factor in decisions such as: a) Replacing incarceration with alternative penalties for low-risk individuals, b) Assessing whether an offender is suitable for community supervision programs, c) Informing probation conditions and the level of supervision required in each case.

- *Key findings:*

Risk assessment tools like COMPAS are becoming increasingly prevalent. In the U.S., more than 60 such tools are in use at various stages of the criminal justice process. While some, such as those in Virginia and Pennsylvania, are developed by state governments, most are owned by private companies.

From a legal perspective, the lack of transparency poses significant challenges to the right of defense. Access to the algorithm was denied to protect Northpointe's intellectual property, the company behind COMPAS. This creates dual challenges: a technical "black box" due to the algorithm's opacity, and a legal “black box” due to trade secret protections. Together, these barriers make it nearly impossible for affected individuals to fully understand or contest how the system operates, undermining the ability to mount an adequate legal defense[14].

It is highly problematic for judicial systems to rely on tools that defendants cannot examine, as this contradicts the fundamental right to a fair defense[15]. Defending against an algorithm is extremely difficult if its inner workings are unknown.

One of the most contentious criticisms of these tools is their potential to introduce bias, thereby violating the principle of equality. In this case, concerns were raised that COMPAS factored gender into its risk assessment, introducing bias. The Wisconsin Court dismissed this claim. However, a subsequent investigation by ProPublica revealed racial biases in COMPAS. The study found that the software was more likely to assign higher risk scores to black defendants than to white ones, sparking an ongoing discussion about fairness and bias in algorithmic decision-making[16].

2.3.3. CORRECTIONAL SERVICE: psychological and actuarial risk Assessment Tools, Canada

Date of final decision: June 13, 2018
Authority: Federal Court of Appeal of Canada (Cour d’appel fédéral)

- *Origin of the case:*

The case originates from Jeffrey G. Ewert, an Indigenous Métis inmate serving two concurrent life sentences, who challenged the use of psychological tests, referred to as assessment tools or actuarial tests, to assess the risk of criminal recidivism and to assess psychopathy in inmates by the CORRECTIONAL SERVICE of Canada (CSC). Ewert argued that these tools were developed and tested primarily on non-Indigenous populations, and their validity when applied to Indigenous offenders had not been established through empirical research. He claimed that this reliance on the tools violated section 24(1) of the Corrections and Conditional Release Act (CCRA), which requires that any information used

[14] LIU, H., LIN, C. and CHEN, Y., “Beyond State v Loomis: artificial intelligence, government algorithmization and accountability”, *International journal of law and information technology* 27, nº 2, 2019, p. 135, available at: https://ssrn.com/abstract=3313916.

[15] NIEVA FENOLL, J., *Inteligencia artificial y proceso judicial*, Marcial Pons, 2018, p. 140.

[16] ANGWIN, J. LARSON, J. MATTU, S., KIRCHNER, L.,“There’s software used across the country to predict future criminals. And it’s biased against blacks.”, *ProPublica,* 23 May 2016, available at:
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
Northpointe's response to the report can be found at: https://www.equivant.com/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity/

by the CSC be as accurate, up-to-date, and complete as possible. Additionally, he argued that the CSC's use of these tools infringed on his rights under sections 7 and 15 of the Canadian Charter of Rights and Freedoms.

- *Decision:*

The Supreme Court of Canada found that the CSC had been using psychological and actuarial tools that were primarily developed and tested on non-Indigenous populations, potentially leading to cultural bias or cross-cultural variance when applied to Indigenous offenders. This presented a risk of systemic discrimination against Indigenous prisoners, as the tools could overestimate the risk of recidivism or lead to other inaccuracies in the assessment of Indigenous offenders' rehabilitation needs.

For that reason, Justice Wagner formally declared that the CSC breached its statutory obligation under section 24(1) of the Corrections and Conditional Release Act by continuing to rely on these tools without verifying their accuracy when applied to Indigenous offenders. However, the court did not find a violation of Ewert's rights under sections 7 or 15 of the Canadian Charter of Rights and Freedoms.

- *Key findings:*

This case focuses on the growing gap between how Indigenous and non-Indigenous offenders are treated in the criminal justice system. For example, Indigenous offenders are more likely to be classified at a higher security level and less likely to get early release. The gap is due in part to policies that may look neutral on the surface but actually discriminate against Indigenous offenders. That's why it was important to ensure the tools worked, regarding that "the CSC had long been aware of concerns regarding the possibility of these tools exhibiting cultural bias yet took no action to confirm their validity and continued to use them in respect of Indigenous offenders, despite the fact that research would have been feasible".

The judgement emphasized that the CSC must ensure that its policies and practices are responsive to the specific needs and circumstances of Indigenous offenders, acknowledging that "those groups are among the most vulnerable to discrimination in the correctional system".

2.3.4. DYNAMIC TRAFFIC CONTROLS: policing methods, the Netherlands

Date of final decision: November 1, 2016
Authority: Supreme Court of the Netherlands (Hoge Raad)

- *Origin of the case:*

In June 2013, Police in Amsterdam stopped a BMW X6 under a procedure known as "DYNAMIC TRAFFIC CONTROLS". This approach selects vehicles for inspection based on risk factors, including potential associations with criminal activity, rather than traffic law violations alone. During the inspection, to which the driver consented, the Police found a bag containing 993 grams of cannabis, leading to the driver's arrest.

The appellant contested the legality of the selection criteria, arguing that they lacked a legal basis under the Road Traffic Act. The Court of Appeal ruled that the police had used their powers under the Traffic Act arbitrarily, amounting to abuse of authority. This decision was appealed to the Supreme Court, which also considered the risk of racial discrimination in the application of "DYNAMIC TRAFFIC CONTROLS", although no explicit claim of racial bias was raised by the appellant.

- *Decision:*

The Supreme Court reversed the appellate court's decision, holding that the use of police powers under the Traffic Act is lawful as long as it maintains a connection to traffic enforcement objectives. The Court clarified that secondary motives, such as suspecting the involvement of individuals in criminal activity, do not invalidate an inspection, provided there is a legitimate traffic-related reason for the stop, such as checking the driver's license or vehicle documentation. In this case, the inspection was deemed lawful.

- *Key findings*:

While the case did not involve algorithmic or AI-driven systems, it raised critical questions about profiling and the potential for discriminatory practices. The Court asserted that police actions must not rely solely on characteristics such as ethnicity or religion, as this would constitute unlawful discrimination subject to legal challenge.

This judgment sets a precedent for addressing AI profiling techniques and ensures that future methods used in traffic controls or similar procedures are scrutinized for fairness and lawfulness, providing safeguards against potential abuse, particularly for vulnerable groups.

2.3.5. CrimSAFE: automated screening in housing applications, United States

Date of final decision: August 7, 2020
Authority: United States District Court for the District of Connecticut

- *Origin of the case:*

Carmen Arroyo applied for housing on behalf of her son, Mikhail Arroyo, at a complex managed by WinnResidential. The application was denied after a CrimSAFE report identified “disqualifying records”, citing a withdrawn charge for retail theft. This charge, which occurred before Mikhail’s disabling accident, never resulted in a conviction.

Carmen Arroyo and the Connecticut Fair Housing Center (CFHC) filed a lawsuit against CoreLogic Rental Property Solutions, LLC (RPS), arguing that CrimSAFE’s design violated the Fair Housing Act (FHA) by disproportionately affecting applicants based on race, national origin, and disability. While CrimSAFE does not issue automated decisions, it flags potentially disqualifying records, heavily influencing housing providers’ decisions. The plaintiffs contended that the tool's operation and the lack of transparency regarding flagged records led to discriminatory outcomes.

- *Decision:*

The Court ruled on a summary judgment motion, partially siding with both parties. It found significant evidence that CrimSAFE’s use had a disparate impact on Latino and African-American applicants, citing systemic biases reflected in U.S. arrest and incarceration statistics.

The Court also held that RPS’s policy of withholding details about flagged criminal records impeded fair housing access for individuals protected under the FHA. Additionally, the lack of transparency and clarity surrounding flagged records was deemed potentially unfair under the Connecticut Unfair Trade Practice Act (CUTPA).

- *Key findings:*

This case underscores the risks posed by automated tools like CrimSAFE in housing decisions. The judgment highlighted how a lack of transparency and inadequate human oversight can perpetuate systemic discrimination, particularly against racial minorities and individuals with disabilities.

The judgment emphasizes the necessity for fairness, transparency, and accountability in automated decision-making systems, especially in housing, where the stakes for vulnerable groups are significant. It also sets a precedent for scrutinizing similar tools for their indirect discriminatory impacts.

2.4. Women and persons exposed to gender-based discrimination

2.4.1. VIOGÉN: Risk Assessment Tools for Recidivism, Spain

Date of final decision: September 30, 2020
Authority: Spanish National Court (Audiencia Nacional)

- *Origin of the case:*

The Comprehensive Monitoring System in cases of Gender-based Violence (VIOGÉN System) is a tool created by the Spanish Ministry of the Interior and operational since 2007. Its main objective is to assess the risk of victims facing future aggression and, based on this assessment, to implement appropriate protective measures. The system assigns a risk level (non-appreciated, low, medium, high, or extreme), and police officers can adjust the automatic result based on additional information.

In this judgement, the Spanish National Court considered a case in which a victim requested a protection order that was denied. The VIOGÉN system classified the risk as “non-appreciated”, and despite clear indications that warranted further investigation, the authorities maintained this risk level. This assessment influenced the judge’s decision to deny the protection order, and the victim was tragically murdered a month later.

- *Decision:*

The Spanish state was found liable for the inadequate protection provided by the police, as they failed to accurately assess the real risk to the victim and left her unprotected.

- *Key findings:*

Although VIOGÉN does not currently use AI (though it is expected to do so in the future)[17], it illustrates the risk of over-reliance on the results generated by a high-risk AI system. The dangerous automatism seen in this case is echoed in the EU AI Act, specifically Article 14.4.b, which warns against automation bias[18].

[17] Secretariat of State for Security of the Ministry of Spain, press release, 15 December 2020, available at: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2020/151220-inteligencia.aspx

[18] According to research by González-Álvarez et al., 95% of police officers chose not to change the risk score suggested by VioGén González Álvarez, J.L., López Ossorio, J.J., Urruela, C. and Rodríguez Díaz, M., Integral Monitoring System in Cases of Gender Violence VioGén System, Behavior & Law Journal, vol. 4, nº 1, 2018, available at: https://behaviorandlawjournal.com/BLJ/article/view/56/65

Given this high level of trust in VIOGÉN's risk assessments, it is crucial to conduct a thorough evaluation of the system's operations, which has not yet been done[19]. This lack of transparency raises concerns about potential biases in completing the form, algorithmic biases within the system, and, most importantly, its accuracy in predicting risk. The results not only influence police measures but also judicial decisions, affecting the rights of both victims and perpetrators, a serious violation of the right to effective judicial protection under Article 24 of the Spanish Constitution and, in extreme cases, the right to life.

2.5. Children and adolescents

2.5.1. DEEPFAKES: generative artificial intelligence, Chile

Date of final decision: August 9, 2024
Authority: Court of Appeals of Santiago, Chile (Corte de Apelaciones de Santiago de Chile)

- *Origin of the case:*

The case originated when parents of female students at Saint George's College filed a constitutional protection action against the school. They argued that the school failed to apply appropriate sanctions against male students who used AI-based tools (DEEPFAKES) to create and distribute explicit images of their daughters.

The Public Prosecutor's Office initiated an *ex officio* investigation and we will have to wait for this judgment on the possible commission of a crime.

- *Decision:*

The Court ruled in favor of the plaintiffs, determining that the school acted illegally and arbitrarily by not applying adequate disciplinary measures as outlined in the school's regulations. The court ordered the expulsion of the responsible students for the 2025 academic year.

- *Key findings:*

[19] MARTÍNEZ GARAY, L. (coord), *Three predictive policing approaches in Spain: VioGén, RisCanvi and Veripol. Assessment from a human rights perspective*, 2022, available at: perspective: https://regulation.blogs.uv.es/files/2024/05/Three-predictive-policing-perspectives-web-17.06.24.pdf and ETICAS FOUNDATION, *Adversarial Audit of the VioGén System*, 2022, available at:
https://eticasfoundation.org/wp-content/uploads/2024/09/Eticas_Audit_of_VioGen.pdf

The court identified that the use of DEEPFAKES by the students to create explicit images of minors constituted a serious violation of the girls' rights. The AI-generated content exploited vulnerable groups (underage students), affecting their psychological and moral well-being, and illustrating the potential for AI misuse in educational and social contexts.

2.5.2. PARCOURSUP: algorithmic transparency in higher education admissions, France

Date of final decision: April 3, 2020
Authority: Constitutional Council of France (Conseil Constitutionnel)

- *Origin of the case:*

The National Union of Students of France (UNEF) challenged the use of algorithms in France's higher education admissions process, specifically the "PARCOURSUP" platform. They argued that Article L. 612-3 of the Education Code restricted access to information about the criteria and methods used by higher education institutions to assess applications. This lack of transparency, according to the UNEF, violated the right to access administrative documents as guaranteed by Article 15 of the Declaration of the Rights of Man and of the Citizen of 1789.

UNEF also contended that the opacity of the algorithms hindered effective judicial protection for rejected students. Without access to detailed information about the selection criteria, students found it difficult to appeal decisions or demonstrate whether their rejection was based on arbitrary grounds or violated their rights.

- *Decision:*

The Constitutional Council upheld the constitutionality of Article L. 612-3 but imposed conditions to ensure transparency and safeguard fundamental rights. The Court ruled that final admissions decisions cannot rely solely on algorithms; they must include human evaluation by application review committees and validation by the institution's director.

The Council mandated transparency by ensuring that candidates are entitled to general information about the qualifications and criteria required for admission to each program. Additionally, rejected candidates have the right to request specific details about the criteria and pedagogical reasons applied in their individual cases.

The judgment also required higher education institutions to publish a post-admissions report outlining the selection criteria and the role of algorithms, ensuring privacy protections for candidates.

- *Key findings:*

This decision reinforces the importance of transparency and access to information in algorithmic decision-making, especially in contexts like education where significant rights are at stake. By prohibiting fully automated decisions without human oversight, the judgment mitigates risks of discrimination and arbitrary administrative actions.

The judgment establishes clear standards for the use of algorithms by public institutions, particularly in sensitive areas affecting vulnerable groups. It balances academic freedom with the rights of applicants, ensuring that algorithmic tools are subject to robust transparency obligations and human oversight to protect fairness in admissions processes.

2.5.3. CHARACTER.AI: generative AI and product liability, United States

Date of final decision: Pending resolution
Authority: United States District Court for the Middle District of Florida, Orlando Division

- *Origin of the case:*

Megan Garcia, representing the estate of her deceased son Sewell Setzer III, filed a lawsuit against Character Technologies, Inc. (Character.AI), Google LLC, Alphabet Inc., and the co-founders of Character.AI. The complaint alleges that the generative AI product “CHARACTER AI” was defectively designed, causing psychological harm that led to Sewell’s self-harm and eventual suicide.

CHARACTER AI is accused of recklessly creating and marketing an AI product with anthropomorphic “characters” that allegedly deceived and emotionally manipulated the 14-year-old. The lawsuit claims that these AI characters fostered an emotional dependence on Sewell, simulating human-like relationships that contributed to his mental distress. The allegations include violations of the Florida Deceptive and Unfair Trade Practices Act (FDUTPA), strict liability for defective products, and negligence, asserting that Character.AI failed to adequately warn about the psychological risks associated with its use, particularly for minors.

- *Decision:*

The case remains pending and awaits judicial resolution. The complaint alleges that CHARACTER AI’s design included dark patterns and intentional anthropomorphic elements that encouraged emotional reliance, particularly among adolescents. It further claims that the application lacked appropriate safeguards, such as parental controls and age verification, increasing the psychological vulnerability of minor users.

- *Key findings:*

This case illustrates the potential dangers of generative AI products designed without robust safety measures, particularly for vulnerable groups like minors. The allegations suggest that the anthropomorphic qualities of AI characters in “CHARACTER AI” created intimate and potentially harmful interactions, including promoting behaviors detrimental to mental health.

The lawsuit emphasizes the need for transparency, oversight, and safety mechanisms in generative AI technologies to mitigate risks, particularly for minors who may struggle to differentiate between reality and fiction in immersive AI environments. If successful, this case could set significant precedents for regulating the design and promotion of AI products to safeguard mental health and prevent harm to vulnerable users.

2.5.4. TIKTOK: addictive design and exploitation of youth mental health, United States

Date of final decision: October 8, 2024
Authority: Superior Court of California, County of Santa Clara

- *Origin of the case*:

This case originated from a lawsuit filed by the State of California, represented by Attorney General Rob Bonta, against TikTok Inc. and its related entities, including ByteDance. The lawsuit alleges that TIKTOK intentionally designed its platform to be addictive, prioritizing extended user engagement and advertising revenue at the expense of users' mental health and well-being, especially among young people. Specific features, such as auto-play videos, infinite scrolling, and a personalized recommendation system, are accused of exploiting psychological vulnerabilities.

The lawsuit also highlights TIKTOK’s alleged collection of personal data from minors without verifiable parental consent, violating both state and federal laws, including the Children’s Online Privacy Protection Act (COPPA). Internal documents revealed that TIKTOK was aware of these risks, which include anxiety, depression, sleep disorders, and even suicidal ideation among young users, yet failed to take meaningful actions to address them. Additionally, TIKTOK's recommendation system, designed to maximize engagement through unpredictable rewards, exploits the neuroplasticity of young, developing brains. Beauty filters and intrusive notifications are identified as elements that reinforce compulsive behavior and amplify risks such as body image issues. The state argues that TIKTOK knowingly prioritized profits over user safety, with misleading safety tools and a lack of meaningful protections for minors.

- *Decision*:

At this preliminary stage, the court has not yet issued a final decision. California is seeking injunctive relief to halt TIKTOK's harmful practices while the litigation continues. Proposed measures include suspending features that promote compulsive use, restricting beauty filters linked to body image concerns, and ceasing the collection of minors' personal data without parental consent.

The lawsuit centers on potential violations of California's unfair competition and false advertising laws. If these violations are proven, possible sanctions include fines for each infraction, restitution of unlawful profits, and structural reforms to safeguard young users. The complaint also alleges that TIKTOK misled the public about the effectiveness of its safety features, such as restricted mode and screen time limits, which are easily bypassed and fail to provide real protection. The outcome could set precedents for regulating technology platforms, particularly regarding data privacy, competition, and the ethical use of AI.

- *Key findings*:

The case underscores the risks of unregulated AI systems in exploiting vulnerable groups, particularly children and adolescents. TIKTOK's AI-driven recommendation system, described in the lawsuit as a tool for maximizing engagement through constant feedback loops, is central to the lawsuit. These features exploit developing brains, increasing the risk of addictive behavior, anxiety, depression, and even suicidal ideation.

The lawsuit brings to light that TIKTOK was aware of the mental health risks posed by its platform but failed to implement meaningful safeguards. Features like beauty filters, driven by AI, promote unrealistic beauty standards, exacerbating body dissatisfaction and self-esteem issues among teenagers. Despite marketing safety tools such as restricted mode and screen time limits, internal evidence revealed that these features were both ineffective and easily circumvented.

The case emphasizes the urgent need for enforceable regulations governing AI in platforms targeting young users. Such frameworks should include algorithmic transparency, robust human oversight, and enhanced protections for minors. Without these measures, AI systems designed for engagement and profit can have profound, long-lasting impacts on young users' mental health and well-being.

2.6. Older persons

2.6.1. EEOC: algorithmic age bias in hiring, United States

Date of Final Decision: August 9, 2023
Authority: United States District Court for the Eastern District of New York

- *Origin of the case:*

In May 2022, the EEOC filed a lawsuit under the Age Discrimination in Employment Act (ADEA), accusing the defendants of developing an algorithm that automatically rejected job applications from candidates above a certain age. Specifically, the algorithm excluded female applicants over the age of 55 and male applicants over the age of 60, affecting over 200 qualified applicants who applied in March and April 2020. These actions, carried out in a programmed and automated manner, created an algorithmic barrier against older applicants, contravening the legal provisions against age discrimination set out in the ADEA.

The defendants, while denying any intentional or illegal discrimination, argued that the tutors did not qualify as employees under the ADEA's definition but rather as independent contractors, a common defence in such cases aimed at circumventing anti-discrimination laws. However, the consent decree signed in August 2023 constitutes an implicit acceptance of the legal obligations imposed by the EEOC, alongside a series of monitoring and compensatory measures.

- *Decision:*

The case was resolved through a Consent Decree that required the defendants to take several corrective actions. As part of the agreement, the defendants committed to paying a total of $365,000 in damages to job applicants who were allegedly affected by discriminatory practices. This monetary compensation aimed to address the harm caused by the unlawful hiring policies.

In addition, the Consent Decree explicitly prohibited the defendants from engaging in hiring practices based on age or sex. They were also barred from collecting birth data before extending a job offer, ensuring compliance with anti-discrimination laws.

The defendants were further obligated to develop and implement comprehensive anti-discrimination policies and establish clear procedures for handling complaints. They were also tasked with providing training to their employees and contractors on U.S. laws prohibiting discrimination, demonstrating a commitment to fostering a fair workplace.

To ensure compliance with these measures, a monitoring mechanism was put in place. This mechanism allowed the EEOC to periodically review the defendants' adherence to the Decree and to seek judicial intervention if violations were identified.

- *Key findings:*

The case illustrates the potential for algorithms to carry out discriminatory actions, in this instance by automatically excluding certain candidates due to age-related factors. These types of algorithms represent a significant legal risk, by encoding exclusionary criteria within hiring processes, they can violate fundamental rights and result in systematic discrimination, often without adequate oversight. The use of such technological tools raises substantial legal challenges concerning the need for oversight and regulation of hiring algorithms, particularly within the tech industry, where their use is increasingly prevalent.

The deliberate configuration of algorithms to discriminate against certain age groups demonstrates how design decisions within AI systems can bear legal consequences. This case highlights the responsibility of corporations to ensure that algorithms comply with equality and non-discrimination regulations. The resolution reinforces that, irrespective of stated intent, the discriminatory outcomes of an automated system can result in legal sanctions and the imposition of structural and monitoring changes.

The defendants argued that the tutors were not employees but rather independent contractors, an attempt to exclude these workers from ADEA coverage. However, this case underscores how anti-discrimination protections may apply to flexible or online business models where the employment relationship is ambiguous or intentionally structured to evade labour responsibilities. This is an emerging issue in employment law and could impact other sectors employing selection algorithms, particularly within the digital economy and gig platforms.

The implementation of monitoring measures by the EEOC sets a precedent regarding the importance of transparency and oversight in enforcing anti-discrimination policies. In this context, the defendants must proactively demonstrate compliance and submit to a regular auditing system. The obligation to communicate clear policies and procedures to all participants in the selection process ensures greater transparency and accountability, crucial elements to prevent algorithmic discrimination and protect the rights of workers and applicants.

The Decree establishes specific training requirements for employees involved in hiring processes, emphasising the need for companies to educate their workforce on the regulatory framework concerning age and gender discrimination. This measure not only seeks to prevent the recurrence of discriminatory incidents but also reinforces the importance of a corporate culture that is mindful of labour rights and equality obligations.

2.7. People with physical, mental, sensory or intellectual disabilities

2.7.1. WORKDAY: Algorithmic Screening Tools in Employment, United States

Date of final decision: July 12, 2024
Authority: United States District Court for the Northern District of California

- *Origin of the case:*

Derek Mobley filed a lawsuit against Workday, Inc., alleging that its algorithmic applicant screening tools discriminate against African-Americans, individuals over the age of forty, and those with disabilities. Mobley, who is African-American, over forty, and has mental health conditions, claimed that WORKDAY's artificial intelligence system automatically screened out over 100 of his job applications, despite his qualifications and experience. Many of these applications were rejected within minutes, indicating that the system operates without meaningful human oversight.

Mr. Mobley argued that the bias stemmed from flaws in the algorithm's training data and the use of cognitive and personality tests, which he claimed disproportionately impacted protected groups. He contended that these practices violated multiple anti-discrimination laws, including Title VII of the Civil Rights Act and the Americans with Disabilities Act.

- *Decision:*

The court issued a mixed judgment on WORKDAY's motion to dismiss. It concluded that Workday acts as an agent for its employer clients, as its algorithmic tools hold decision-making authority over candidate selection. The court found it plausible that WORKDAY's system generates a disparate impact against African-Americans, individuals over forty, and people with disabilities, citing Mobley's zero success rate in the screening process and the alleged biases in the system's training data and tools like pymetrics and personality tests.

However, the court dismissed Mobley's claim of intentional discrimination, ruling that awareness of a tool's discriminatory effects does not equate to intent. The court acknowledged Mobley's argument that WORKDAY was aware of these adverse impacts but found no evidence that WORKDAY deliberately intended to discriminate.

- *Key findings:*

This case demonstrates the potential for AI systems in recruitment to perpetuate biases against protected groups, such as racial minorities, older candidates, and individuals with disabilities. The judgment underscores the liability of hiring platforms like WORKDAY for the outcomes of their automated systems, especially when these systems wield significant influence over hiring decisions.

While the case will proceed to trial, the court's recognition of the plausible disparate impact of algorithmic hiring tools serves as a significant precedent. It emphasizes the need for transparency, accountability, and mitigation of biases in AI-driven employment processes to protect vulnerable groups from systemic discrimination.

2.8. People with chronic illnesses or health conditions that lead to discrimination

2.8.1. ChatGPT: Generative Artificial Intelligence, Colombia

Date of final decision: August 2, 2024
Authority: Constitutional Court of Colombia (Second Chamber of Review)

- *Origin of the case:*

The Second Review Chamber considered a constitutional protection (tutela) action filed by Blanca, the mother of a minor diagnosed with autism spectrum disorder (ASD), against a Health Promotion Entity (EPS). She sought the protection of her son's fundamental rights to health and dignified life, due to the EPS's refusal to (i) exempt her son from co-payments and moderating fees, (ii) cover transportation costs for him to attend his therapy sessions, and (iii) guarantee comprehensive treatment.

The second-instance judge used ChatGPT 3.5 to inquire about certain legal questions regarding the fundamental right to health for minors diagnosed with ASD and incorporated the questions and answers into the reasoning of the judgment. The Constitutional Court analyzed whether the use of AI in judicial reasoning could constitute a violation of due process by effectively replacing the judge's decision-making role.

- *Decision:*

The Court concluded that due process had not been violated, as the judicial decision was made before the consultation with ChatGPT, and the questions and answers were transcribed afterwards. Therefore, the validity of the court's decision was not in question, as it had been made before using ChatGPT.

While the Colombian Court does not prohibit the use of AI systems, it urges judges to use them in a way that respects fundamental rights, in particular, due process and ensures the independence of the judiciary. To this end, the Court states that judicial officers and employees must adhere to the principles of (i) transparency, (ii) accountability, (iii) privacy, (iv) non-substitution of human reasoning, (v) seriousness and verification, (vi) risk prevention, (vii) equality and fairness, (viii) human oversight, (ix) ethical regulation, (x) adherence to best practices and collective standards, (xi) continuous monitoring and adaptation, and (xii) competence.

In addition, the court orders the dissemination of a guide, manual or guidelines on the implementation of generative AI in the judiciary, in particular, on the use of ChatGPT.

- *Key findings:*

Although the Chamber confirms that due process was not violated in this case, it does not dismiss the potential risks of hallucinations, discriminatory bias and other risks associated with AI. Specifically, it notes that ChatGPT's results "may be biased because this tool generates outputs by generalising from the knowledge used in its training data when confronted with new inputs, which may perpetuate the biases present in its training data, producing responses based on stereotypes or favouring certain groups or ideas. This poses a significant risk, for example, to minority populations or persons with special constitutional protections" (par. 260).

2.8.2. ADULT BUDGET CALCULATION TOOL, United States

Date of final decision: June 11, 2015
Authority: United States Court of Appeals for the Ninth Circuit

- *Origin of the case:*

The plaintiffs, representing a class of disabled individuals, challenged the way the Idaho Department of Health and Welfare (IDHW) calculated budgets for home and community-based services (HCBS). The plaintiffs argued that the system used to allocate services was unclear, arbitrary and often reduced necessary services without adequate explanation, violating their due process rights under the Fourteenth Amendment.

- *Decision:*

The Ninth Circuit ruled that Idaho's Medicaid budgeting system violated due process by failing to provide adequate notice or procedural protections to Medicaid recipients when their service budgets were adjusted or reduced.

- *Key findings:*

The Court emphasised that vulnerable groups, particularly Medicaid recipients with disabilities, require enhanced procedural protections. The lack of clear information and justification for budget cuts using the ADULT BUDGET CALCULATION TOOL, which automatically determined participants' service budgets based on personal information, could significantly harm this group, stressing the need for transparency and accountability in administrative decisions affecting their essential care and services.

3. Potential risks for vulnerable groups

This section presents AI systems that impact the general population but may pose indirect or potential risks to vulnerable groups, exacerbating their vulnerability and limiting their capacity to respond. The analysis focuses primarily on two scenarios: the widespread use of facial recognition tools and the application of AI systems within the justice system.

3.1. Facial recognition technologies

3.1.1. Facial recognition technology, Spain

Date of final decision: July 27, 2021
Authority: Spanish Data Protection Agency

- *Origin of the case*:

Mercadona, a Spanish distribution company, decided to implement an early detection system using facial recognition technology (FRT) in its stores due to the risk associated with criminal acts. This decision was prompted by the high number of crimes committed in its centres across Spain, which pose a risk to both customers and employees, as well as to the company's goods. The data processing includes the capture, matching, storage and destruction in case of negative identification of the biometric image captured of any person entering the supermarket.

In addition, facial recognition consists of comparing a dubious biometric sample, obtained from one or more images of a person, against a database of biometric samples already associated with the identity of a person, which have been previously registered through one or more photographs. For this purpose, "the questionable biometric samples" are transformed into patterns. Subsequently, through facial recognition, the biometric samples are compared with the previously stored indubitable template, through algorithmic calculations that are evaluated based on previously established matching thresholds.

This system is used for cases of (i) final convictions resulting from criminal proceedings in which Mercadona is a party to the proceedings, and the images are obtained from the video-surveillance cameras located in its facilities that were provided as evidence in the proceedings; (ii) convictions in which Mercadona is not a party to the proceedings, in the case of restraining orders for crimes committed against its employees and the Courts and Tribunals directly request Mercadona's collaboration in relation to the scope of the restraining order to the victim's workplace, in order to enforce the restraining orders. And to carry out this purpose it will process the data (i) of a convicted person; (ii) Mercadona's potential customers; and, (iii) Mercadona's employees.

- *Decision*:

The case raises several violations in relation to the (i) processing of special categories of data; (ii) principle of minimisation of personal data; (iii) principle of data protection by design; (iv) impact assessment, considering that it is not carried out in an adequate manner. And, on the basis of the GDPR, the sanction is graduated taking into account the size of the company and the type and volume of data processed, the categories of data subjects, such as minors and vulnerable persons, as well as the scope, since it is carried out remotely, massively and indiscriminately. In addition to taking into account that the data controller did not carry out the prior consultation, the security system will carry out a systematic and exhaustive evaluation of personal aspects of natural persons on a large scale of special category data.

- *Key findings:*

The processing can be considered extremely high-risk and unacceptable, as it could lead to massive and indiscriminate surveillance. It involves automatically capturing biometric data using pre-set algorithms that analyze the processed image of each individual, potentially deriving sensitive information such as race, gender, emotional state, illnesses, genetic defects, substance consumption, and more. This violates the principle of data minimization under the GDPR.

3.1.2. Facial recognition technology, Russia

Date of final decision: October 04, 2023
Authority: European Court of Human Rights (Third Section)

- *Origin of the case*:

The case stems from the use of facial recognition technology (FRT) to identify and arrest Nikolay Glukhin for his participation in a peaceful protest, where he held a cardboard figure bearing a political message. The FRT system was used to locate him in the metro, gather evidence against him and arrest him on the basis of the Public Acts Act for failing to give prior notice of the demonstration. NiKolay Glukhin was convicted of an administrative offence under applicable Russian law. The use of the FRT system involved the processing of biometric data of the person concerned without his consent.

- *Decision*:

The European Court of Human Rights (ECtHR) concluded that the use of facial recognition in this context violated several principles of the European Convention on Human Rights: (i) right to private and family life, in this case, the processing of biometric data using FRT did not meet the requirements of lawfulness and necessity. The applicable Russian law was vague and lacked adequate safeguards that did not guarantee protection against abuse and arbitrariness; (ii) freedom of expression, in relation to the conviction of Glukhin for his participation in a peaceful protest associated with the use of FRT, as it has an intimidating effect on the exercise of this fundamental right; (iii) right to a fair trial, the absence of clear safeguards on the use of FRT and its misuse to gather evidence in an administrative proceeding highlighted the lack of transparency and judicial review that compromised the process.

- *Key findings:*

The indiscriminate use of this technology can result in mass surveillance with a deterrent effect on freedom of expression and assembly. This disproportionately affects activists and political minorities, who may face reprisals for exercising fundamental rights. It also poses risks of abuse and arbitrary use of biometric data, eroding public trust in technology and government institutions. The ECtHR emphasised the need to establish a robust legal framework limiting the use of AI in public contexts to protect human rights and prevent state abuses.

3.1.3. Facial recognition technology, United States (State of New Jersey)

Date of final decision: June 07, 2023
Authority: Superior Court of New Jersey, Appellate Division

- *Origin of the case*:

In this case, facial recognition technology (FRT) was used as the basis for the robbery charge. Following a robbery, police sent real-time images of the suspect to the New York Police Department Crime Centre, which identified Mr. Arteaga as a possible suspect. On the basis of this identification coupled with the identification of witnesses on photographs, Mr. Arteaga was charged. The defence argued that it should have access to information about the FRT including the algorithm, the error rate and the database used in order to assess the credibility of the system and challenge the identification made. It was indicated that the accuracy of this FRT is critical to Mr. Arteaga's defence rights as unreliability could lead to a case of mistaken identity.

- *Decision*:

The New Jersey Court of Appeal reversed the decision and decided to grant the access requested by Mr. Arteaga’s defence, indicating that the lack of transparency in the use of FRT could compromise the right to a fair trial. The violations it identified were based on the (i) right to a fair trial with adequate due process, and the lack of transparency that led to the incrimination of Mr. Arteaga deprived the defence of the right to a fair trial. Arteaga deprived the defence of the necessary means to challenge the reliability of the system and, consequently, to have a fair trial; (ii) the right to confrontation, determining that the defence has the right to be able to examine the FRT, including the source code and its parameters due to the novelty of the technology associated with the accuracy in determining the validity of the identification; (iii) principles of fairness and reliability, the court indicated that the defence should be able to assess whether the FRT introduced racial bias or technical flaws.

The court ordered that information be provided to enable a proper analysis and assessment of this technology in court. The information to be provided included all other documents relating to the operation and accuracy of the FRT, as these elements are essential to ensure the defence and, if necessary, to be able to challenge the identification and question the investigation carried out.

- *Key findings:*

This case shows how the use of AI-based technologies in judicial processes without due guarantees of transparency and adequate supervision can compromise the fundamental rights of individuals, especially minorities. The lack of transparency and the possibility of error or bias resulting in inaccurate facial identification errors, coupled with the lack of access to information about their operation, makes it necessary to design and implement adequate controls over the use of AI-based technologies in the judicial system. Without adequate controls, discrimination and vulnerability may be exacerbated, as algorithms are often less accurate with ethnic minorities, increasing the risk of misidentification in these groups. Without access to information about the operation of FRT, those affected may be placed at a disadvantage, increasing the risk of unfair decisions, therefore reinforcing the importance of accountability in the use of AI-based technologies.

3.1.4 Facial recognition technology, United States (California)

Date of final decision: August 08, 2019
Authority: United States Court of Appeals for the Ninth District

- *Origin of the case*:

In this case, a group of Facebook users in Illinois sued the organisation for using facial recognition technology (FRT) without consent on the basis of a violation of the Illinois Biometric Information Privacy Act (BIPA). Facebook used the technology to create facial templates from photographs uploaded by users in breach of the duty to disclose and obtain their written consent, invading their right to privacy. The lawsuit focused on the applicable privacy regulations, which require a public retention policy and obtaining informed consent before collecting and storing users' biometric data.

- *Decision*:

The Appellate Court ruled in favour of the plaintiffs and confirmed that Facebook violated the Illinois BIPA by detailing several breaches (i) of the requirements for prior consent and information due to Facebook's failure to inform users of the processing of their biometric data, nor obtain prior consent, violating users' right to privacy by collecting biometric identifiers without their knowledge or consent; (ii) lack of data retention and destruction policies by failing to establish a public retention and destruction plan for biometric data, increasing the risk of misuse of personal information and planning doubts about how long this sensitive data could be stored; (iii) infringement of the right to privacy by creating and storing facial templates without authorisation, compromising users' right to privacy.

The Court found that these acts constituted a “concrete harm” under the US Constitution, allowing the plaintiffs to bring the lawsuit. Additionally, the court found that the violations of the Illinois BIPA pose a significant privacy risk justifying certification of the class action on behalf of those affected in Illinois. While the Ninth Circuit did not rule on the merits of the case, this case is the first certified BIPA class action.

- *Key findings:*

We can draw from this case the importance of the risks associated with AI technology in the collection of biometric data without the guarantees of transparency and consent of the right holder. Unauthorised collection of biometric data can seriously affect vulnerable individuals or groups who may not be in a position to object or fully understand the scope of AI-based technologies. Lack of effective regulation on biometric data protection and opacity or lack of transparency in the operation of AI use can lead to massive privacy breaches, exposing individuals to risks of surveillance and misuse of their data without their knowledge or control.

3.1.5 CLEARVIEW AI, Canada

Date of final decision: December 14, 2021
Authority: the Office of the Privacy Commissioner of Canada, the Quebec Information Access Commission (the Commission d’accès à l’information du Québec), the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta

- *Origin of the case*:

The case arises out of an investigation being conducted jointly by several Canadian data protection authorities. The investigation is triggered by the use of AI-based facial recognition technology (FRT) used to collect without the consent of the data subjects more than 3 billion images. Clearview AI's goal was to make facial images obtained from public online sources (social media and other websites) available to law enforcement for the purpose of identifying individuals. Clearview justified its position and actions on the grounds that the images were publicly available and did not require consent.

- *Decision*:

The Canadian authorities concluded that Clearview AI violated privacy laws by collecting, using, and disseminating personal data without prior consent or a legitimate basis for processing. The authorities noted that the mass collection of facial images for the purpose of creating a readily recognisable database is tantamount to mass surveillance, which is inappropriate as it does not meet the standards of proportionality to privacy that are part of the expectations of Canadian citizens.

The authorities determined that Clearview AI had to cease its services in Canada by ceasing to market its facial recognition tool, and delete personal data and stop collecting, using and disseminating biometric data of Canadian citizens and delete all data already collected. It was established that Clearview had not obtained consent for the processing of biometric data and had failed to adequately inform data subjects about such activities.

Clearview's processing of biometric data was deemed to be a form of “mass surveillance” impacting on the privacy and well-being of Canadian citizens, including minors, rejecting Clearview's arguments that it violated applicable privacy laws. During the investigation, Clearview withdrew from the Canadian market expressing its disagreement with the findings and recommendations without committing to compliance with the authorities’ directives.

The decision stresses the personal data protection stance on the massive and indiscriminate use of biometric data and the necessary compliance for its processing in accordance with applicable personal data protection regulations.

- *Key findings:*

It highlights the risks associated with the use of AI-based technologies in facial recognition systems, especially if it involves the massive and indiscriminate processing of biometric data without adequate control. The resolution illustrates how these technologies without supervision or consent can lead to mass surveillance and affect the right to privacy of the subjects concerned. It also exposes individuals to risks of privacy violations and misidentification, especially in specific demographic groups, which could lead to discrimination. It is important that those responsible for developing AI-based technology put in place measures to protect the fundamental rights of individuals.

3.1.6 CLEARVIEW AI, Australia

Date of final decision: July 07, 2021
Authority: Office of the Australian Information Commissioner (OAIC).

- *Origin of the case*:

The Office of the Australian Information Commissioner (OAIC) decided to launch an investigation into Clearview AI for the mass collection of facial images (more than 3 billion) of individuals without informing or obtaining their consent. Clearview carried out mass scraping of data from the internet, including social media data, and stored it in a database. These facts motivated the investigation as they violated applicable privacy laws and, consequently, the legality of using this data to identify individuals without their consent or being informed.

- *Decision*:

The OAIC found that Clearview AI breached several legal provisions of the applicable privacy legislation (i) on the collection of images and biometric data without prior consent, in breach of the principle of lawful and fair collection and processing of personal data; (ii) on the lawfulness of the processing of personal data, and on the means used for the collection of images such as "data scraping" without adequately informing the biometric data subjects; (iii) on the duty of information and transparency in the collection of personal data by failing to take reasonable measures to inform the data subjects of the images; (iv) on the relationship with the principle of accuracy of information by failing to ensure that the information used was accurate, up-to-date and relevant as the images collected were not verified, nor the veracity of the data processed.

- *Key findings: the role of AI in vulnerable groups:*

This case study highlights the danger inherent in the processing of biometric data without a legitimate basis using AI-based technologies, with a particular impact on vulnerable groups such as minors or people without sufficient resources to protect their privacy. The indiscriminate use of facial recognition technologies that also generate databases to be exploited without obtaining the consent of the data subjects may entail an interference with a high impact on the privacy of natural persons. Moreover, the lack of transparency in their use increases the risk of discrimination, mass surveillance and manipulation, disproportionately affecting those who cannot understand or do not have control over the use of their personal data.

3.1.7 CLEARVIEW AI, France

Date of final decision: November 26, 2021
Authority: National Commission for Information Technology and Liberties (Commission nationale de l'informatique et des libertés (CNIL)).

- *Origin of the case*:

Clearview AI uses proprietary technology that indexes images of faces from social networks, professional web sites, blogs and other web pages, without discriminating between images of adults and minors. It collects images and uses software to generate unique facial prints from these images. With a database of more than 10 billion images, the company offers a search service where users can upload a photo and the system identifies similar faces in its database and displays the images found, along with links to the pages where they appear.

Clearview AI defends its service as a police tool that aims to identify potential perpetrators and victims of crime using only a photograph.

The National Commission for Information Technology and Liberties (CNIL) received several complaints in 2020 about the problems faced by personal data subjects when trying to exercise their rights of access and deletion of personal data with Cleraview.

- *Decision*:

The CNIL has assessed the performance of Clearview AI's facial recognition software, which processes extracted photographs to generate detailed profiles of individuals. This software stores not only images, but also metadata, such as URLs of the source websites, which allows inferring patterns of behaviour, interests and locations, especially when they come from social networks or media. This data processing, being massive and continuous, makes it possible to track individuals over time, which is considered a form of “behavioural tracking”.

The CNIL has concluded that Clearview AI collects biometric data, such as facial recognition, without a basis that legitimises the data processing, in breach of the applicable regulations. It considers that the legitimising basis of the legitimate interest of the company does not justify this type of data processing, given its particularly intrusive nature on the privacy of individuals. In addition, a violation of users' access and other rights has been identified, as the company limits the exercise of these rights, allowing only two access requests per year and providing information only for the last twelve months.

Although the CNIL did not impose a financial fine, it has ordered Clearview AI to cease processing the personal data of French citizens within two months, to ensure the full exercise of data subjects' rights, especially the right of access, and to delete personal data for which there is no legitimate basis for processing. The company has also been requested to demonstrate compliance with these measures.

- *Key findings:*

The risk of harm to data subjects is significant in this case. The police authority has processed personal data without legal considerations and without any control or assessment of the intrusiveness of the processing on the personal integrity of individuals. The improper and unlawful use of these biometric data may involve forms of discrimination that could lead to the automatic categorisation of individuals, with corresponding biases.

3.1.8 CLEARVIEW AI, United States (Illinois)

Date of final decision: July 25, 2022
Authority: United States District Court for the Northern District of Illinois, Eastern Division

- *Origin of the case*:

In Clearview AI, Inc, Consumer Privacy Litigation, the plaintiffs alleged that Clearview AI processed their personal data without consent. The company collected more than 3 billion images of faces obtained from social media and other websites. They used these images to create unique biometric identifiers using AI algorithms, in violation of the Illinois Biometric Information Privacy Act (BIPA). Under BIPA, the processing of biometric data requires the informed consent of the data subject, which did not occur in this case. The plaintiffs argued that these practices violated their privacy and exposed their personal data to potential abuse.

- *Decision*:

The Northern District Court of Illinois found that Clearview AI had violated several provisions of BIPA, (i) by failing to implement a policy specifying the retention period and procedures designed and implemented for the deletion of biometric data; (ii) by collecting biometric data without notifying data subjects in advance or obtaining their written consent, in breach of the obligation to disclose the purpose and duration of data processing; (iii) it was questioned whether it used biometric data for commercial purposes, which is prohibited without explicit authorisation. Clearview IA sold access to its biometric database to third parties, such as government agencies and private companies, without complying with the legal restrictions.

The court denied Clearview IA’s motion to dismiss the case for lack of jurisdiction, reaffirming the applicability of BIPA to protect Illinois citizens, and declared that Clearview IA's practices constituted an invasion of privacy in processing the biometric data.

- *Key findings:*

This case illustrates how the unregulated use of AI to collect and analyse biometric data can put vulnerable groups at risk, including minorities and individuals who are unaware of the misuse of their data. The absence of consent and the monetisation of this data can increase the risks of mass surveillance, discrimination and infringement of fundamental rights. In addition, these practices may have a deterrent effect on freedom of expression and participation in public activities, as individuals may fear that their biometric data will be processed without authorisation.

3.1.9 CLEARVIEW AI, Netherlands, Greece, Hamburg, Italy.

Date of final decision: Netherlands, November 26, 2021; Greece, July 13, 2022; Hamburg, February 12, 2021; Italy, February 10, 2022
Authority: Nederland Personal Data Authority (Autoriteit Persoonsgegevens, AP); Hellenic Data Protection Authority [Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα); Bundesbeauftragter für den Datenschutz und die Informationsfreiheit (BfDI)(Hamburg Commissioner for Data Protection and Freedom of Information); Italian Data Protection Authority (Garante per la Protezione dei Dati Personali).

- *Origin of the case*:

Clearview AI, a technology company that uses facial recognition through web scraping techniques, collected images from social media and other sources to create a massive database of human faces, which were then processed by AI to identify individuals. The images were collected without the consent of the image holders and therefore breached the legality, information and transparency obligations imposed by the General Protection Data Regulation (GDPR). Through its “Clearview for law-enforcement and public defenders” service, it processed images of individuals without an adequate basis for legitimisation and without providing the necessary information to those concerned.

- *Decision*:

Clearview was found to be in breach of several provisions of the GDPR (i) it processed personal data without a valid legitimate basis. It did not obtain the consent of the data subjects, nor did It demonstrate a legitimate interest sufficient to justify the processing; (ii) the company processed biometric data, a special category of personal data, without the explicit consent of the data subjects; (iii) it failed to provide adequate information to the data subjects, in breach of the obligation to inform them about the processing of their data; (iv) it failed to respond to data subjects' requests for access, which prevented data subjects from exercising their right of access to their data; (v) it failed to appoint a representative in the European Union, which also constitutes an infringement.

Clearview AI was sanctioned by several data protection authorities in Europe (German, Dutch, Italian, and Netherlands) for the unlawful processing of personal data. The sanctions included administrative fines and orders for data deletion. AI

Data Protection Authorities ordered Clearview AI to immediately stop processing personal data of individuals located in their respective countries and in the European Union; and it was required to delete all images and biometric data that were obtained from individuals without their consent. It was also ordered to appoint a representative in the European Union, due to its processing of personal data of European individuals.

In addition to the injunction, Clearview AI was ordered to comply with other regulatory requirements set out in the GDPR, including ensuring transparency in its data processing practices and the processing of biometric data.

- *Key findings:*

Mass data collection without consent and lack of transparency are particularly dangerous for vulnerable groups, such as minors or people who are unaware of how their personal data is processed. In addition, the indiscriminate use of AI to collect, store and process biometric data can lead to serious privacy violations, impact discrimination and be misused by governments or security agencies, highlighting inequalities and risks to citizens' fundamental rights. The lack of measures to delete individuals' data that are no longer publicly available further amplifies concerns about the control and protection of personal data.

3.1.10 CLEARVIEW AI, Sweden

Date of final decision: February 10, 2021
Authority: Sweden Authority for Privacy Protection (Integritetsskyddsmyndigheten (IMY))

- *Origin of the case:*

Clearview AI, a facial recognition application provided by a US company, allows users to upload an image, which is then biometrically compared with a large database of images collected from the internet. When it became known that the Swedish Police Authority had used this application, concerns about its legality led the authority's data protection officer to advise the National Forensic Centre and the National Operational Department to clarify that such use was prohibited.

The applicable regulations on Police Processing within the Scope of the Data Protection Act (PBDL) states that the Police Authority is responsible for the processing of personal data carried out by the authority. This means that, it is the obligation of the Police Authority to ensure that all data processing carried out by the authority has, among other matters, a legitimate basis, a lawful purpose and with appropriate technical and organisational measures adequate and appropriate to the risk of the processing of these personal data. In each individual case, it should be assessed which measures are necessary, taking into account, inter alia, which personal data are being processed. The Police Authority did not provide (i) any policy for the processing of personal data by employees; (ii) any documentation on how employees will be trained; (iii) how the internal procedure will be implemented in the organisation; (iv) no training or equivalent activity was carried out.

Following this, the Sweden Authority for Privacy Protection (IMY) launched an investigation, asking the Police Authority to clarify whether it had used Clearview AI and the legal basis for processing the data. The Police Authority confirmed that some employees had used the application in several ongoing criminal investigations. It was found that biometric data, in the form of individuals' facial images, had been processed in connection with these cases, but there was no legal assessment or documentation on how this data was handled.

Biometric data is classified as sensitive personal data under privacy regulations, and its processing is only allowed in certain cases. IMY's investigation raised concerns that using Clearview AI, which involves matching individuals' biometric data with large amounts of unfiltered personal data collected from the internet, likely does not meet the strict necessity requirements outlined in the Criminal Data Act and the Criminal Data Directive.

- *Decision*:

The IMY launched an investigation into the Swedish Police Authority after discovering that several employees had used the Clearview AI tool without the necessary authorization to identify individuals. The processing of biometric data through facial recognition violated the Criminal Data Act, as the authority failed to conduct a required data protection impact assessment and implement organizational measures to ensure compliance with data protection regulations.

IMY emphasized the risks associated with using third-party technology from a foreign country to process sensitive biometric data. The Swedish Police Authority did not demonstrate that the processing was absolutely necessary for its intended purpose, nor did it clarify what happened to the data entered into Clearview AI. IMY ordered the authority to implement training and organizational measures to ensure compliance, notify affected individuals by September 15, 2021, and ensure that personal data entered into Clearview AI was deleted.

The investigation found aggravating circumstances, including prolonged use of Clearview AI, access to a large volume of personal data without transparency, and the processing of sensitive biometric information for facial recognition. Despite these factors, the number of affected individuals was relatively small, which was considered a mitigating factor.

As a result, IMY fined the Swedish Police Authority SEK 2,500,000 (approximately 250,000€) for breaching the Data Protection Act. This penalty highlights the importance of adhering to legal frameworks when processing personal data, particularly sensitive biometric data.

- *Key findings:*

The data entered into the application by the Police Authority has been privacy-sensitive, and it remains unclear what has happened to this personal data after its entry. IMY considers the risk and harm to data subjects in this case to be significant. The Police Authority processed personal data without legal justification or an assessment of its impact on individuals' privacy. The improper and unlawful use of biometric data raises concerns about potential discrimination and the automatic categorization of individuals, introducing inherent biases.

3.1.11 CLEARVIEW AI, United Kingdom

Date of final decision: May 18, 2022
Authority: Information Commisoner´s Office (ICO)

- *Origin of the case:*

Clearview AI has developed an image search engine that allows its customers, such as law enforcement, to compare a "probe image" (an image of interest) against an extensive database of images, metadata and URLs, collectively known as the "Clearview database".

To perform this search, the client provides Clearview AI with a probe image, from which Clearview generates a unique face vector and compares it to similar vectors in its database. This process returns a list of thumbnail images with direct links to where each image appears on the Internet. Clearview does not provide any explicit identification or attribute analysis for the probe image; instead, the client is responsible for reviewing the URLs and determining the identity, attributes, location or behaviour of individuals in the search results. The images in the Clearview database have been "scraped" from internet access sources, including social media, without filtering out those showing UK residents.

- *Decision*:

The ICO found that Clearview AI breached applicable personal data protection legislation and details multiple breaches of the GDPR in its processing of personal data, as well as its impact on the rights of UK residents. In particular, (i) by failing to adequately inform data subjects about the collection and processing of their public images, a practice that is not in line with the expectations of the individuals concerned; (ii) it lacks a data retention policy, keeping the images in its database indefinitely, which does not comply with the principle of limited retention; (iii) lacks a legitimate basis for processing the data, and processes special category personal data without a legitimate basis; (iv) hinders the exercise of the rights as users have to provide their own photograph for verification, which discourages the exercise of these rights, and does not allow users to request the deletion of their personal data; (v) failed to carry out an impact assessment.

The ICO considers these facts to be serious and has decided to issue an Enforcement Notice to compel Clearview AI to comply with the provisions of the UK GDPR. Clearview AI's response denied the applicability of this regulation and did not propose alternatives to comply.

The ICO proposed a fine of £17 million or 4% of the company's total annual worldwide turnover and ordered the company to delete all personal data of UK residents within specified timeframes. It also required the company to cease collecting data from UK residents, stop matching images against its database, and conduct a DPIA to be submitted to the ICO.

- *Key findings:*

The risk of harm to data subjects is high in this case. The police authority has processed personal data without legal considerations and without any control or assessment of the intrusiveness of the processing on the personal integrity of individuals. The improper and unlawful use of these biometric data may involve forms of discrimination that could lead to the automatic categorisation of individuals, with corresponding biases.

3.2. AI systems within the justice system

In this section, two types of cases are analyzed: one concerning the analysis of DNA evidence and another involving the use of videoconferencing in the context of criminal proceedings. Additionally, two AI systems are discussed: one designed to assess profiles and facilitate border controls, and another aimed at identifying risk patterns and enabling preventive actions by police authorities.

3.2.1. TrueAllele: DNA evidence, United States (California)

Date of final decision: January 9, 2015
Authority: California Court of Appeal, Second Appellate District, Division Four

- *Origin of the case:*

The case revolves around the prosecution of Martell Chubbs, accused of the murder of Shelley H. in 1977. In 2011, DNA evidence generated a genetic profile matching that of Chubbs, leading to his arrest and prosecution in 2012. The prosecution utilised the TrueAllele software, developed by Cybergenetics, to conduct probabilistic DNA analysis. This analysis concluded that the likelihood of a match between the evidence and Chubbs was extraordinarily high.

The core issue arose when the defence sought access to the source code of the TrueAllele software, arguing that this analysis was the sole evidentiary link to Chubbs. The defence contended that, without the source code, it could not adequately assess the reliability of the software. Cybergenetics refused to disclose the source code, invoking the trade secret privilege, which led to a legal conflict between the rights of the defence and the protection of intellectual property.

- *Decision:*

The trial court initially ordered the disclosure of the source code, citing the accused's constitutional right to confront the evidence and witnesses presented against him. However, the Court of Appeal overturned this decision, holding that the defence had failed to demonstrate a particularised need that would justify breaching the trade secret privilege. Consequently, the Court upheld Cybergenetics' commercial interests, determining that the information already provided by the company was sufficient for the defence to understand the methodology and reliability of TrueAllele.

- *Key findings:*

In the case at bar, the Court of Appeal grappled with the intersection of fundamental constitutional rights and commercial interests, presenting a nuanced legal challenge. At the core of the dispute was a tension between the defendant's right to a fair trial, specifically, the right to confront and challenge the evidence against them, and the proprietary rights safeguarding intellectual property, as enshrined under California Evidence Code, Section 1060. The court acknowledged the defendant's right to a fair trial, which encompasses the ability to confront and scrutinize evidence, including underlying methodologies. However, it clarified that this right is not absolute and does not extend to unrestricted access to protected information absent a compelling justification. In this case, the defendant's demand to access the source code of the TrueAllele software, a proprietary algorithm central to the forensic evidence, was critically examined against these principles. The court underscored that such access must be substantiated by demonstrating a specific and compelling need, particularly in the pre-trial phase, where the balance of rights is especially delicate.

The defence contended that the validity of TrueAllele's forensic analysis could not be adequately assessed without examining its source code. They argued that the software's reliance on assumptions might undermine its conclusions, necessitating transparency for effective cross-examination. However, the court found this argument speculative, noting the absence of concrete evidence showing how access to the source code would materially affect the evidentiary reliability or the defence's ability to contest the case. The court highlighted that Cybergenetics, the software's creator, had provided extensive materials, including operational manuals, peer-reviewed articles, and supervised access to the software. These measures were deemed sufficient to enable a meaningful evaluation of the software's reliability without compromising its proprietary integrity.

The court's decision firmly upheld the classification of the TrueAllele source code as a trade secret under California Evidence Code, Section 1060. It recognized the substantial and irreparable financial harm that could result from disclosing the source code, which embodies a competitive advantage in a highly specialized and competitive field. In doing so, the court affirmed the importance of protecting technological innovation and commercial interests, particularly within advanced domains such as artificial intelligence. This case serves as a landmark precedent, reflecting the increasing reliance on artificial intelligence and algorithms in judicial processes, particularly for forensic evidence analysis. While such technologies offer unparalleled precision and efficiency, their inherent opacity poses challenges to ensuring transparency and safeguarding the right to an adequate defence. The judgment strikes a delicate balance, allowing the use of proprietary algorithms in evidence while safeguarding procedural rights through alternative validation mechanisms. Central to the court's reasoning was the extensive scientific validation of the TrueAllele system. Peer-reviewed studies and its widespread application in judicial proceedings across jurisdictions were cited as compelling evidence of its reliability. The court concluded that this body of validation obviates the need for direct access to the source code, ensuring confidence in the software's results without undermining its proprietary protections.

3.2.2. iBorderCtrl: AI and transparency in border controls, European Union

Date of final decision: December 15, 2021
Authority: General Court of the European Union

- *Origin of the case:*

The claimant, Patrick Breyer, sought access to documents concerning the "iBorderCtrl" project, which employs artificial intelligence to assess profiles and facilitate border controls. The Research Executive Agency (REA) partially denied access, relying on exceptions outlined in Regulation (EC) No. 1049/2001. The claimant subsequently filed an action seeking the annulment of this decision, alleging violations of his right to access information and a restrictive interpretation of transparency rules.

- *Decision:*

The General Court determined that the REA failed to conduct an exhaustive review of the claimant's initial request, thereby breaching the objectives of Regulation No. 1049/2001. While recognising the legitimacy of protecting commercial interests, the Court required a more detailed and less restrictive assessment of the requested information. Furthermore, it reaffirmed that exceptions must be applied with strict proportionality.

- *Key findings:*

In addressing the implications of AI on fundamental rights, it becomes evident that the integration of disruptive technologies must be carefully balanced with the preservation of privacy, transparency, and core individual liberties. The consideration of ethical assessments and risk profiling plays a pivotal role in this equilibrium, serving as essential mechanisms to ensure the protection of individual rights amidst technological innovation. The principle of transparency is a cornerstone of democratic governance, and exceptions to public access must be stringently interpreted. The Court, in this instance, reaffirmed that restrictions on access to information should be invoked only in instances of absolute necessity. Legitimate interests, such as the safeguarding of commercial operations, must not become a pretext for obstructing public oversight of technologies with far-reaching societal implications. Transparency, particularly regarding AI systems, remains indispensable to maintaining public trust and accountability. A further tension arises in the interplay between confidentiality and the public interest. In the arguments presented by the REA, the protection of intellectual property rights and commercial strategies was positioned as a justification for withholding information. However, the Court pointed out that, in scenarios involving an overriding public interest, such as the deployment and assessment of artificial intelligence in sensitive or impactful contexts, transparency takes precedence. This prioritization ensures that the public has the opportunity to engage in an informed discourse on matters that significantly shape societal frameworks.

3.2.3 TrueAllele: DNA and algorithmic, United States (New Jersey)

Date of final decision: February 3, 2021
Authority: Appellate Division of the Superior Court of New Jersey

- *Origin of the case:*

This case originates from an incident that occurred in 2017, in which the defendant was linked to a fatal shooting through DNA evidence processed by TrueAllele. The defence challenged the reliability of this software and requested access to its source code to evaluate whether it correctly implemented the underlying scientific methods. The trial court denied this request, citing the need to protect the trade secrets of the developer, Cybergenetics.

- *Decision:*

The appellate court held that the refusal to provide access to the source code compromised the defendant's rights to a full defence and due process. The court argued that an independent review of the software under a protective order was essential to assess its reliability, particularly given the significant technical complexity and potential margin of error inherent in probabilistic algorithms.

- *Key findings:*

The court, in addressing procedural rights and access to evidence, underscored the fundamental importance of allowing the defence to scrutinize the source code of TrueAllele. Such access is indispensable for ensuring a rigorous evaluation of the tool's reliability, which directly impacts the defendant's constitutional rights to due process and the ability to confront and challenge the evidence presented by the prosecution. Without this access, the defendant's right to a fair trial risks being undermined by the opacity of the technological processes employed against them.

When examining the complexity of probabilistic algorithms, the court acknowledged that systems like TrueAllele interpret intricate DNA data through the application of mathematical models and coding logic. While advanced, these tools are not immune to human error in their design and implementation. The court highlighted previous issues with software such as STRmix and FST, which revealed vulnerabilities only uncovered through independent audits. Such findings emphasize the necessity of thorough examination and validation of these technologies to prevent miscarriages of justice. The court also grappled with the tension between protecting trade secrets and upholding justice. While TrueAllele's source code is commercially sensitive, it cannot be shielded at the expense of a defendant's right to a fair trial. To reconcile these competing interests, the court proposed a model of limited access under a protective order, allowing necessary judicial scrutiny while safeguarding the proprietary interests of the developer. Drawing on legal precedents, the court referenced cases such as State v. Chun, where access to the source code of a breathalyser device was mandated to ensure fairness. The decision also reflected on challenges encountered in jurisdictions like New York, where independent reviews of similar forensic software uncovered critical errors. These precedents reinforce the principle that technological tools used in criminal proceedings must withstand rigorous examination.

3.2.4 hessenDATA: police automated tools, Germany

Date of final decision: February 16, 2023
Authority: Federal Constitutional Court of Germany (Bundesverfassungsgericht)

- *Origin of the case:*

The case arose from provisions in the Public Safety and Order Act of the State of Hesse (§ 25a HSOG) and the Police Data Processing Act of the State of Hamburg (§ 49 HmbPolDVG), which allow police authorities to use automated tools to analyse large volumes of data, identify risk patterns, and take preventive measures. The claimants argued that these provisions violated their fundamental rights, particularly the right to informational self-determination (Article 2(1) in conjunction with Article 1(1) of the Basic Law) and other privacy-related rights. They contended that the collection, processing, and analysis of data through advanced technologies, without adequate safeguards, represented a disproportionate interference with their constitutional rights.

- *Decision:*

The Federal Constitutional Court declared certain specific provisions of both laws unconstitutional, as they did not meet the necessary thresholds to justify the preventive use of automated data analysis tools. In particular, the Court emphasized the need for stricter and more specific limitations on data use and the implementation of safeguards to ensure the protection of fundamental rights. While recognizing the importance of automated technologies for the prevention of serious crimes, the Court underscored that such technologies cannot be implemented without a clear and strict legal framework that adheres to the principle of proportionality.

- *Key findings:*

The Court addressed the significant implications of automated data processing on the right to informational self-determination, particularly through the use of platforms such as hessenDATA in Hesse and similar systems in Hamburg. It determined that these tools constitute an interference with this fundamental right by enabling the generation of novel insights from previously unrelated data sets. The algorithms employed in these systems facilitate the creation of detailed profiles and the identification of patterns or correlations that would otherwise remain undetectable through manual methods, thereby heightening the risks associated with the use of personal data. Central to the Court's reasoning was the principle of purpose limitation, which mandates that personal data may only be utilized for the specific purposes for which it was originally collected. Any subsequent use involving a change in purpose must be explicitly justified by a separate legal basis. In this context, the Court affirmed the principle of proportionality, stressing that automated data analysis may have far-reaching and intrusive effects that extend beyond those of mere data collection. Such effects demand careful scrutiny to ensure that individual rights are not disproportionately affected.

The Court further recognized the risks associated with discrimination and profiling in the application of artificial intelligence and algorithms. It noted that these systems, particularly when employed for electronic profiling or predictive policing, can exacerbate the risk of generating unfounded suspicions. This is especially concerning when statistical correlations are used to assess individuals who are not involved in criminal activities, potentially influencing operational decisions by law enforcement without adequate human oversight or accountability. In addressing the constitutional framework governing the use of such automated technologies, the Court concluded that serious interferences with the right to informational self-determination must be subject to strict conditions akin to those regulating covert surveillance measures[20]. These conditions include the requirement for clear and specific legal justification, as well as the establishment of robust safeguards to protect against misuse. Transparency, legal protection, and administrative oversight were identified as essential elements to ensure that the deployment of these technologies complies with constitutional standards and does not undermine the fundamental rights of individuals.

3.2.5 Videoconferencing in Criminal Proceedings: Remote Trials and Due Process, Chile

Date of final decisions: December 10, 2020; March 30, 2021; March 31, 2021; March 31, 2021
Authority: Constitutional Court of Chile (Tribunal Constitucional de Chile)

[20] COTINO HUESO, L., "Una regulación legal y de calidad para los análisis automatizados de datos o con inteligencia artificial. Los altos estándares que exigen el Tribunal Constitucional alemán y otros tribunales, que no se cumplen ni de lejos en España", *Revista General de Derecho Administrativo*, nº 63, 2023, p. 8.

- *Origin of the cases:*

These cases arose from challenges to Law No. 21226, which allowed the use of videoconferencing for judicial hearings and oral trials in criminal matters during the COVID-19 pandemic. The plaintiffs, defendants in criminal proceedings held in pretrial detention, argued that the virtual format compromised their constitutional rights to adequate defense and due process. They contended that videoconferencing limited their ability to prepare their defense and review evidence effectively, thus infringing on fundamental judicial guarantees.

- *Decisions:*

In all cases, the Constitutional Court upheld the constitutionality of Law No. 21226, recognizing the necessity of implementing technological measures in response to the public health emergency. However, the Court underscored that these adaptations must align with constitutional principles, particularly the defendant's right to defense and the critical role of judges in subjectively assessing evidence.

- *Key findings:*

The judgments collectively emphasize the importance of human oversight in judicial proceedings, particularly in the use of technology like artificial intelligence. While the Court acknowledged that technology can support certain procedural aspects, it strongly cautioned against the replacement of human judgment by AI in areas such as evidence assessment.

The decisions warned of the potential risks of automation bias and the erosion of the “Human in the Loop” principle, which is essential for ensuring impartiality and fairness. These cases set a significant precedent by reaffirming that the integration of technology in the judicial process must not compromise constitutional rights, especially for vulnerable individuals. The judgments highlight the need to balance technological efficiency with robust safeguards to maintain the integrity of the criminal justice system.

4. Conclusions

This final section summarizes the main findings of this jurisprudential study, referencing vulnerable groups, the authorities issuing the judgments, the time period in which the decisions were rendered, the country, the plaintiffs, and the principles and rights outlined by the OECD that are impacted.

4.1. Vulnerable groups

We have identified 7 resolutions addressing people in situations of poverty or social exclusion (plus 3 cases very similar to SyRI which, although they have not reached the courts, are relevant to this study); 5 involving persons belonging to racial or ethnic minorities, including migrants, refugees, and indigenous peoples; 4 concerning informal and precarious workers; 4 related to children and adolescents; 2 involving people with chronic illnesses or health conditions that lead to discrimination; and 1 resolution for women and persons exposed to gender-based discrimination, 1 for older persons, and 1 for people with physical, mental, sensory, or intellectual disabilities.

As noted in the introduction, in some cases, more than one vulnerability factor converges. This is a critical consideration when addressing the impacts of AI, highlighting the importance of adopting an intersectional approach. Specifically, in the case of AFR Locate in the United Kingdom, the AI system affects women and ethnic minorities; in CRIMSAFE, in the United States, applicants are impacted based on race, national origin, and disability. Similarly, in WORKDAY, also in the United States, discrimination targets African Americans, individuals over the age of forty, and those with disabilities

A last section introduces AI systems that affect the general population but may create indirect or potential risks for vulnerable groups, increasing their susceptibility and limiting their ability to respond effectively. The analysis focuses on two primary scenarios: the widespread use of facial recognition technology (FRT) and the deployment of AI systems within the justice system.

The first case study examines the situation of citizens who may be randomly impacted by facial recognition systems. The analysis systematically addresses cases such as Mercadona (Spain), Glukhin v. Russia (ECtHR), and the cases in New Jersey and California (United Stated). It also includes some incidents generated by CLEARVIEW AI in Canada, Australia, France, Illinois, the Netherlands, Greece, Hamburg, Italy, Sweden, and the United Kingdom.

The second set of cases explores the potential impact of AI on citizens interacting with criminal proceedings. This includes the analysis of 2 cases involving DNA evidence in the United States, the iBorderCtrl system developed by the European Union, issues concerning automated policing and fundamental rights in Germany, and 4 instances of the use of videoconferencing in criminal proceedings in Chile.

4.2. Courts and authorities deciding the cases

In 35 cases, the judgments were issued by courts, predominantly high courts, as detailed below:

- Supreme Courts: Wisconsin State, the Netherlands
- Constitutional Courts: France (Constitutional Council), Chile (Constitutional Court), Italy (Council of State), Germany (Federal Constitutional Court)
- Courts of Appeal (federal): United States, United Kingdom, Chile, Canada.
- Courts of Appeal (state): New Jersey, California
- National High Court: Spain, Australia

Additionally, we include 1 judgment from the Court of Justice of the European Union (CJEU), 1 from the General Court of the European Union (EGC) and another 1 from the European Court of Human Rights (ECtHR). For its relevance, we have also included 1 Opinion of the Advocate General of the CJEU.

Some cases were decided by lower courts, such as the California Northern District Court, The Hague District Court, Connecticut District Court, Bologna Ordinary Court, Georgia Cobb County State Court, New York District Court, Florida Middle District Court, and the Superior Court of California.

The remaining decisions were issued by data protection authorities, specifically: the Information Commissioner's Office (UK), the Office of the Privacy Commissioner (Canada), the Hamburg Commissioner for Data Protection and Freedom of Information (Germany), the Swedish Authority for Privacy Protection, the Office of the Australian Information Commissioner, the National Commission for Information Technology and Civil Liberties (France), the Spanish Data Protection Agency, the Italian Personal Data Protection Authority, and the Hellenic Data Protection Authority (Greece).

4.3. Time period

The period of the decisions starts in 2013 and ends in 2024, with 2 cases (Bosco in Spain and Workday in the United States) still open and another 2 pending (Character AI and a facial recognition technology case in California), so we will have to wait for a final decision. Most cases are concentrated from 2020 onwards. Some of the decisions in the early years are not strictly AI, but when considering the use of algorithmic systems, it is interesting to study them because of the high impact they have on the decision-making process.

4.4. Geographical scope

In terms of geographical scope, almost all cases have been resolved by authorities in European countries or the United States, with the exception of Australia, Canada and Chile, as well as the judgements of 3 supranational courts.

In Europe, the countries analysed were: the Netherlands, France, Italy, Germany, the United Kingdom, Spain, Switzerland and Greece. In the United States, the cases came from the following states: Wisconsin, New Jersey, California, Connecticut, Georgia, New York and Florida.

4.5. Plaintiffs

This study has identified a variety of plaintiffs[21].

- Class actions:
    - ROBODEBT, Australia
    - ADULT BUDGET CALCULATION TOOL, United States
    - Facial recognition technology, United States

- Individuals:
    - AMAZON, United States
    - iBorderCtrl, European Union
    - TrueAllele, United States (New Jersey)
    - CHARACTER.AI, United States
    - WORKDAY, United States
    - Videoconferencing in Criminal Proceedings, Chile
    - OFQUAL, United Kingdom
    - SCHUFA, Germany
    - COMPAS, United States
    - CORRECTIONAL SERVICE, Canada
    - VIOGÉN, Spain
    - ChatGPT, Colombia
    - DUN & BRADSTREET, Austria
    - hessenDATA, Germany
    - TEACHER ALLOCATION ALGORITHM, Italy

[21] This section does not include Clearview AI cases where there are no plaintiffs as such, but rather complainants, many of whom are not identified.

- Glunkhin vs.Russia, Russia
- Francisco Arteaga, United States (State of New Jersey)

- Multiple plaintiffs:
  - CalWIN, California (legal services organizations, welfare rights organisations, private individuals).
  - SyRI, the Netherlands (civil society interest groups and private individuals).
  - DEEPFAKES, Chile (parents of female students)
  - CrimSAFE, United States (private individual and Connecticut Fair Housing Center)

- NGO:
  - AFR Locate, United Kingdom (a civil liberties campaigner)
  - BOSCO, Spain (Civio Foundation)

- Trade Unions
  - UBER, United Kingdom (Uber drivers in London)
  - DELIVEROO, Italy (Trade unions Filt Cgil, Filcams Cgil, and Nidil Cgil)
  - PARCOURSUP, France (The National Union of Students of France - UNEF)

- Public authorities:
  - TrueAllele, United States (California) (Prosecutor Office)
  - DYNAMIC TRAFFIC CONTROLS, the Netherlands (Advocate General at the Court)
  - TIKTOK, California (Attorney General)
  - EEOC, United States (Equal Employment Opportunity Commission)

4.6. Rights and principles affected

The OECD AI Principles were initially adopted in 2019 and updated in May 2024. The Principles guide AI actors in their efforts to develop trustworthy AI and provide policymakers with recommendations for effective AI policies. These 6 principles are: Inclusive growth, sustainable development and well-being; Respect for the rule of law, human rights and democratic values, including fairness and privacy; Transparency and explainability; Robustness, security and safety; Accountability[22]. What follows is a summary of what the resolutions under review have to say about each of these issues, in some cases, being interconnected in such a way that the protection or violation of one affects the others.

4.6.1 Respect for the rule of law, human rights and democratic values, including fairness and privacy

This principle underscores that AI actors should respect the rule of law, human rights, and democratic, human-centered values throughout the entire AI system lifecycle. This notion is reflected in various judgments, such as the PARCOURSUP case, where the Constitutional Council of France emphasized that final admissions decisions cannot rely solely on algorithms; instead, they must include human evaluation by application review committees and validation by the institution's director. Similarly, the Federal Court of Australia, in the ROBODEBT case, highlighted that governments cannot exclusively rely on automated systems when formulating and implementing public policy.

The significance of this principle lies in the potential violation of other fundamental principles if it is not upheld. This was evident in the CRIMSAFE case, where the judgment underscored that inadequate human oversight can perpetuate systemic discrimination, particularly against racial minorities and individuals with disabilities.

An example of this is the COMPAS case, where the use of an AI system resulted in discrimination against a black individual. State v. Loomis has become one of the most significant and widely publicized judgments, as the court stated that while an AI system cannot be used to determine the severity of a sentence, it can be considered a relevant factor in: a) deciding whether to substitute a prison sentence with an alternative for individuals classified as low risk, b) determining if the offender's risk of recidivism is suitable for participation in supervision programs and community services, and c) informing decisions regarding the terms and conditions of parole, as well as the appropriate supervision and control in each case. This judgment effectively approved the use of risk assessment tools within the justice system.

One of the implications of being human-centered is avoiding over-reliance on the results generated by an AI system, a phenomenon known as automation bias (Article 14.4.b of the EU AI Act). This undesirable effect was precisely what occurred in the VIOGÉN case, where excessive trust in an AI risk assessment

[22] OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449), Adopted on: 22 May 2019; Amended on: 03 May 2024, available: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

system, which indicated a “non-appreciated” risk level, influenced the judge’s decision not to provide protective measures for the victim. Tragically, this decision resulted in the murder of a woman who was a victim of gender-based violence. All the cases concerning videoconferencing in criminal proceedings considers the “Human in the Loop” principle crucial for maintaining impartiality and fairness. Decisions in such cases emphasize the need for human oversight to prevent undue reliance on automated systems, ensuring that critical judgments remain grounded in human evaluation rather than algorithmic outputs. Thus, the SCHUFA case is particularly significant as it marks the first time the CJEU addressed the definition of a “decision based solely on automated processing”. The Court analyzed Article 22 of the GDPR on automated decision-making and adopted a broad interpretation, concluding that a score derived from a probability value constitutes a “fully automated decision” if it meaningfully impacts the decision-making of a third party.

This human-centered principle also encompasses values such as non-discrimination and equality, freedom, dignity, autonomy of individuals, privacy and data protection, diversity, fairness, social justice, and internationally recognised labour rights.

Among these, non-discrimination and equality is the right most frequently addressed in the analyzed cases. The following examples illustrate the potential for algorithms to engage in discriminatory practices: OFQUAL (schools from deprived areas), DELIVEROO (riders engaged in strike actions), AFR Locate women and ethnic minorities), COMPAS (black defendants), CORRECTIONAL SERVICE OF CANADA (indigenous offender), CrimmSAFE (race, national origin, and disability), WORKDAY (African-Americans, individuals over forty, and people with disabilities), EEOC (female applicants over the age of 55 and male applicants over the age of 60).

The requirement to comply with the law and respect human rights encompasses the right to due process, which came under scrutiny in the ChatGPT case ruled on by the Constitutional Court of Colombia. Although the court found that due process had not been violated, it emphasized that the use of AI systems must respect fundamental rights, particularly due process and judicial independence, while also considering the potential risks of hallucinations or discriminatory bias inherent in such technologies. In a similar vein, other cases address the implications for due process, such as those involving facial recognition technology and the use of TrueAllele in the United States (State of New Jersey), and videoconferencing in criminal proceedings in Chile.

Additionally, specific rights have been violated by the use of AI in other cases. For example, in the DEEPFAKE case, the court identified that the use of deepfake technology against female students constituted a serious violation of their rights, affecting their psychological and moral well-being. The TIKTOK case, which also affects young users, underscores the risks of unregulated AI systems in exploiting vulnerable groups and the urgent need for enforceable regulations governing AI.

The Glukhin v. Russia case highlights how the use of facial recognition technology in protests violated fundamental rights, including privacy, freedom of expression, and assembly. The lack of regulation and disproportionate surveillance discouraged democratic participation, emphasizing the need for robust legal frameworks to ensure AI systems comply with principles of legality, necessity, and proportionality.

Internationally recognized labor rights are another set of rights encompassed within this principle. These were addressed in cases such as UBER, which pointed out the urgent need to modernize labor protections in response to the rise of digital platforms and artificial intelligence, and the AMAZON FLEX APP case, which further underscored this necessity.

4.6.2 Transparency and explainability

AI actors should ensure transparency and responsible disclosure by providing context-appropriate and state-of-the-art information to promote a general understanding of AI systems, including their capabilities and limitations. They should also offer clear and accessible details about data sources, factors, processes, or logic behind AI outputs (predictions, content, recommendations, or decisions). This transparency enables affected individuals to comprehend results and challenge adverse outcomes arising from AI systems.

These precautions are exemplified in several analyzed cases. In PARCOURSUP, the decision underscores the importance of robust transparency obligations and access to information in algorithmic decision-making, particularly in sensitive areas like education, where significant rights are at stake.

In BOSCO, the High National Court of Spain highlights the importance of transparency in the use of algorithms by public administrations and emphasizes the need for regulations to ensure such transparency. However, access to the source code of the social electricity subsidy was denied three times, mainly on the basis of cybersecurity claims disguised as intellectual property.

The same emphasis on transparency is evident in the DUN & BRADSTREET case, where the primary issue was the extent of information the company must disclose to ensure transparency and accuracy in its automated credit assessment process. The Advocate General of the CJEU opined that, under Article 15(1)(h) of the General Data Protection Regulation, DUN & BRADSTREET must provide “meaningful” information about the logic of the automated decision-making process. This includes the methodology and criteria used to evaluate creditworthiness but excludes complex technical details, such as complete algorithms, if they qualify as trade secrets or involve the personal data of third parties.

In the TEACHER ALLOCATION ALGORITHM case, the court also underlines the importance of ensuring clear and accessible accountability mechanisms and access to information to avoid or mitigate disproportionate impacts.

Finally, in the SyRI case, the court takes this principle a step further, asserting that a lack of transparency can lead to discriminatory outcomes, particularly against individuals from lower socio-economic backgrounds or immigrant communities. The Court emphasized the necessity of transparency and accessibility in algorithmic systems, particularly those employed by public administrations, and criticized the “black box” nature of SyRI, stressing that its lack of transparency impeded individuals’ ability to understand and challenge decisions affecting them. This opacity was deemed a violation of fundamental rights, including the right to privacy and protection against discrimination. The court underscored that, in the public sector, the use of artificial intelligence must be accompanied by “white box” systems that ensure transparency and allow for external scrutiny to uphold citizens’ rights.

In this regard, this study reveals the uneven outcomes when requesting information about public algorithms, with both positive and negative experiences, with SyRI being one of the cases that demands greater guarantees of transparency. Also in the New Jersey facial recognition case, where the Court clarified that failure to provide access to information could compromise the right to a fair trial and the right to defence. However, other cases, such as BOSCO, fall on the opposite side, prioritizing intellectual property rights.

4.6.3 Robustness, security and safety

The requirement that AI systems should be robust, secure, and safe implies the adoption of necessary measures to ensure that AI systems do not pose risks of causing undue harm or exhibiting undesired behavior. A relevant example, though not yet resolved by the courts, is the case of CHARACTER AI, where an AI designed without robust safety measures allegedly facilitated intimate and potentially harmful interactions, including the promotion of behaviors detrimental to mental health. This issue is particularly concerning for minors, who may struggle to distinguish between reality and fiction in immersive AI environments. If this case proceeds successfully, it could establish significant precedents for regulating the design and promotion of AI products, emphasizing the need to safeguard mental health and prevent harm to vulnerable users.

4.6.4 Accountability

Accountability in AI requires that AI actors are responsible for ensuring the proper functioning of AI systems and adherence to fundamental principles, taking into account their roles, the context, and the state of the art. This includes ensuring traceability, applying systematic risk management, and promoting collaboration and responsible conduct among AI actors.

The need to meet this requirement is evident in several of the cases analyzed. For example, the TEACHER ALLOCATION ALGORITHM, SyRI, and CRIMSAFE cases relate to the transparency requirement, while the Workday case also relates to bias mitigation. The EEOC case underscores the responsibility of corporations to ensure that algorithms comply with equality and non-discrimination regulations. In the hessenDATA case, the importance of accountability in law enforcement decision-making processes is emphasized. Similarly, in the New Jersey facial recognition case, the court stresses the need for accountability mechanisms in the context of facial recognition technologies.

These findings can be highly valuable for associations, NGOs, and organizations working to defend vulnerable groups, as they provide a map of the main groups affected by AI that have either reached judicial instances or received decisions from data protection authorities. It is particularly important to identify who initiated the proceedings, which rights were violated, and the principles that must be upheld when AI systems are employed.

ANNEX. SUMMARY TABLE

Case	Court/Authority	Year	Vulnerable group
CalWIN (Super. Ct. No. 07CS01306)	Court of Appeal of the State of California, Third Appellate District	2013	People in situations of poverty or social exclusion
Adult Budget Calculation Tool (789 F.3d 962 (9th Cir. 2015)	United States Court of Appeals for the Ninth Circuit	2015	People with chronic illnesses or health conditions that lead to discrimination
TrueAllele (A-4207-19T4)	California Court of Appeal, Second Appellate District, Division Four	2015	Citizens in the justice systems
COMPAS (State v. Loomis, 881, N.W.2d 749)	Wisconsin Supreme Court	2016	People in situations of poverty or social exclusion
Dynamic Traffic Controls (ECLI:NL:HR:2016:2454)	Supreme Court of the Netherlands	2016	Persons belonging to racial or ethnic minorities
Correctional Service of Canada (Ewert v. Canada 2018 SCC 30)	Federal Court of Appeal of Canada	2018	People in situations of poverty or social exclusion
Uber (Case No: A2/2017/3467)	Court of Appeal (Civil Division), United Kingdom	2018	Informal and precarious
Teacher allocation algorithm (N. 02270/2019REG.PROV.COLL. N. 04477/2017 REG.RIC.)	Italian Council of State in the Courts (Sixth Section)	2019	Workers

Case	Court/Authority	Year	Vulnerable group
Facial recognition technology, United States (No. 18-15982)	United States Court of Appeals for the Ninth Circuit	2019	Citizens
SyRI (C / 09/550982 / HA ZA 18-388)	The Hague District Court	2020	People in situations of poverty or social exclusion
VioGén (SAN 2350/2020)	Spanish National Court	2020	Women
CrimSAFE (No. 3:18-CV-705 (VLB))	United States District Court for the District of Connecticut	2020	Persons belonging to racial or ethnic minorities
Parcoursup (Decision No. 2020-834 QPC)	Constitutional Council of France	2020	Children and adolescents
AFR Locate ([2020] EWCA Civ 1058)	Court of Appeal (Civil Division) on Appeal from the High Court of Justice Queen’s Bench Division (UK)	2020	Persons belonging to racial or ethnic minorities
Videoconferencing in Criminal Proceedings (Rol No. 8892-2020; Rol Nº 10.118-2021; Rol Nº 10.156-2021; Rol Nº 10.045-2021)	Constitutional Court of Chile	2020 2021	Citizens in the justice systems
Deliveroo (Case 2949/2019)	Ordinary Court, Bologna, Italy	2020	Informal and precarious

Case	Court/Authority	Year	Vulnerable group
Ofqual (IC-70514-H7K5)	Information Commissioner's Office, UK	2021	People in situations of poverty or social exclusion
Robodebt (Prygodicz v Commonwealth of Australia (No 2) [2021] FCA 634)	Federal Court of Australia	2021	People in situations of poverty or social exclusion
Clearview IA Canada	Office of the Privacy Commissioner of Canada, Quebec Information Access Commission British Columbia Information and Privacy Commissioner, Information Privacy Commissioner of Alberta	2021	Citizens
Clearview IA Sweden (Diary number: DI-2020-2719// A126.614/2020)	Authority for Privacy Protection (IMY)	2021	Citizens
Clearview IA Australia (CII20/00006)	Office of the Australian Information Commissioner (OAIC)	2021	Citizens
Clearview IA Hamburg	Hamburg Commissioner for Data Protection and Freedom of Information	2021	Citizens
Clearview IA Netherlands	Nederland Personal Data Authority	2021	Citizens
Clearview IA France (Decisión No. MED-2021-134)	National Commission for Information Technology and Liberties	2021	Citizens

Case	Court/Authority	Year	Vulnerable group
Mercadona (PS/00120/2021)	Spanish Data Protection Agency	2021	Citizens
TrueAllele (466 N.J. Super. 270 (App. Div. 2021) 246 A.3d 279)	Appellate Division of the Superior Court of New Jersey	2021	Citizens
Amazon (Civil Action File No.: 21-A-2303)	Cobb County State Court, Georgia, USA.	2021	Informal and precarious
I-Border Ctrl (No. 1049/2001)	General Court of the European Union	2021	Citizens
Clearview IA Italy (doc. web n. 9751362)	Italian Data Protection Authority	2022	Citizens
Clearview IA United Kingdom	Information Commisoner’s Office	2022	Citizens
Clearview IA Greece (Original No: 1809 Decision 35/2022)	Hellenic Data Protection Authority	2022	Citizens
Clearview IA Illinois (21-cv-0135 (N.D. Ill. Jul. 25, 2022))	United States District Court for the Northern District of Illinois, Eastern Division	2022	Citizens
Schufa (C-634/21)	European Court of Justice	2023	People in situations of poverty or social exclusion
Glukhin v. Russia (Application no. 11519/20)	European Court of Human Rights (Third Section)	2023	People in mass surveillance with a deterrent effect on freedom of expression and assembly

Case	Court/Authority	Year	Vulnerable group
State of New Jersey v. Francisco Arteaga (Docket No. A-3078-21)	Superior Court of New Jersey, Appellate Division	2023	Citizens
hessenDATA (Case: 1 BvR 1547/19 - 1 BvR 2634/20 -)	Federal Constitutional Court of Germany	2023	Citizens
EEOC (Case n.: 1:22-cv-2565--PKC-PK)	United States District Court for the Eastern District of New York	2023	Older persons
Deepfakes (RIT (Rol Interno de Tribunal): Protección-13557-2024)	Court of Appeals of Santiago	2024	Children and adolescents
ChatGPT (Sentencia T-323 de 2024 Referencia: expediente T-9.301.656)	Constitutional Court of Colombia	2024	People with chronic illnesses or health conditions that lead to discrimination
Dun & Bradstreet Austria (Case C-203/22)	Opinion of Advocate General of the CJEU, Mr. Richard de la Tour	2024	People in situations of poverty or social exclusion
Workday (Derek Mobley vs. Workday, Inc) (Case No. 3:23-cv-00770-RFL)	U.S. District Court for the Northern District of California	2024	People with physical, mental, sensory or intellectual disabilities
Character.AI (Garcia v. Character Technologies, Inc., et al.)	United States District Court for the Middle District of Florida, Orlando Division	2024	Children and adolescents

Case	Court/Authority	Year	Vulnerable group
Bosco (SAN 2013/2024)	Spanish National Court	2024	People in situations of poverty or social exclusion
TikTok The People of the State of California, v. TikTok Inc., et al.	Superior Court of California, County of Santa Clara	2024	Children and adolescents