Stanford – Vienna
Transatlantic Technology Law Forum

A joint initiative of
Stanford Law School and the University of Vienna School of Law

TTLF Working Papers

No. 150

The GDPR and the US Algorithmic Accountability Act: Evaluating Their Adequacy in Regulating AI-Based Decision-Making

Bilge Tugce Yazgan

2026

TTLF Working Papers

Editors: Siegfried Fina, Mark Lemley, and Roland Vogl

About the TTLF Working Papers

TTLF's Working Paper Series presents original research on technology, and business-related law and policy issues of the European Union and the US. The objective of TTLF's Working Paper Series is to share "work in progress". The authors of the papers are solely responsible for the content of their contributions and may use the citation standards of their home country. The TTLF Working Papers can be found at http://ttlf.stanford.edu. Please also visit this website to learn more about TTLF's mission and activities.

If you should have any questions regarding the TTLF's Working Paper Series, please contact Vienna Law Professor Siegfried Fina, Stanford Law Professor Mark Lemley or Stanford LST Executive Director Roland Vogl at the

Stanford-Vienna Transatlantic Technology Law Forum
http://ttlf.stanford.edu

Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305-8610

University of Vienna School of Law
Department of Business Law
Schottenbastei 10-16
1010 Vienna, Austria

About the Author

Bilge Tugce Yazgan is a Turkish attorney and a qualified member of the Istanbul Bar Association. She holds a Bachelor of Laws degree from Bilkent University, Turkey (2021) and an LL.M. in European and International Business Law from the University of Vienna, Austria (2025). Her research interests include data protection, AI regulation, intellectual property law, and international regulatory frameworks.

General Note about the Content

Suggested Citation

This TTLF Working Paper should be cited as:
Bilge Tuğçe Yazğan, The GDPR and the US Algorithmic Accountability Act: Evaluating Their Adequacy in Regulating AI-Based Decision-Making, Stanford-Vienna TTLF Working Paper No. 150, http://ttlf.stanford.edu.

Copyright

Abstract

Artificial intelligence (AI) is rapidly reshaping social, economic, and legal life, raising urgent questions about accountability, fairness, and protecting fundamental rights. While the European Union's General Data Protection Regulation (GDPR) and the newly adopted Artificial Intelligence Act (AI Act) constitute the most ambitious attempts worldwide to regulate AI, their adequacy in practice remains contested. Across the Atlantic, the United States has proposed the Algorithmic Accountability Act (AAA) as a procedural mechanism for algorithmic oversight, yet its uncertain legislative fate and fragmented state-level rules reveal significant governance gaps.

This thesis critically examines the extent to which these frameworks provide adequate safeguards for AI-powered decision-making. It analyzes the doctrinal foundations of the GDPR, the AI Act's risk-based obligations, and the AAA's impact-assessment model, situating them within broader ethical principles such as transparency, fairness, accountability, and human dignity. The study highlights both convergences and divergences through a comparative analysis of EU and U.S. approaches, supported by case studies on recruitment algorithms, credit scoring, and public-sector surveillance. It shows that while the GDPR and AI Act enshrine strong rights and prohibitions, ambiguities in provisions such as Article 22 GDPR and AI Act enforcement uncertainty limit their effectiveness. Conversely, the AAA offers a promising model of procedural accountability but suffers from a narrow scope, weak enforcement, and legislative fragility.

The thesis argues that existing frameworks remain only partially adequate: They address key risks but leave persistent gaps in enforcement, redress, and protection for vulnerable groups. To close these gaps, it recommends harmonizing definitions of "high-risk" and "automated decision systems," mandating stronger transparency obligations (including data disclosure), embedding stakeholder participation, and fostering international regulatory cooperation. By situating its analysis within the broader debate on digital humanism, the thesis underscores that effective AI governance requires robust legal frameworks and ethical commitments to fairness, autonomy, and human dignity.

Keywords

Artificial Intelligence, GDPR, EU AI Act, Algorithmic Accountability Act, Automated Decision-Making, Digital Humanism, Transparency, Accountability, Fairness

Table of Contents

1. Introduction

1.1 AI's Ubiquity and Societal Impact

Artificial Intelligence (AI) has quickly become integrated into nearly every aspect of modern life. From social media feeds and digital assistants to predictive analytics in healthcare and finance, AI-driven systems now inform countless daily decisions. AI is already having a major impact on society and raising pressing questions about how, where, and by whom these impacts will be felt. These systems, powered by large-scale data analytics, increasingly generate inferences that may significantly affect individuals' rights and opportunities, often in contexts where those individuals lack awareness, control, or recourse[1]. On the other hand, the benefits of AI are undeniable. It improved efficiency, data-driven insights, and automation of complex tasks, yet its widespread adoption without any control also brings significant societal risks. Incorrect or biased algorithmic decisions can produce tangible harms, for instance, wrongful arrests traced to flawed facial recognition or online content algorithms that reinforce prejudices, and such failures harm public trust and create concerns that companies deploying AI wield excessive power over individual lives[2]. In short, as AI systems increasingly mediate social, economic, and even legal outcomes, their societal impact has moved to the forefront of public and academic discourse. Ensuring that this impact is beneficial and fair is a central challenge of our time.

1.2 Opacity and Accountability Challenges

[1] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI' (2019) 20 Columbia Business Law Review 494, 494–495.
[2] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press 2018) 1–11.

A defining concern in AI governance is the opacity of many advanced AI models, and that problem comes from the nature of the "black box" of the AI. Complex machine learning systems often operate in ways not readily interpretable by humans, even by their developers. This lack of transparency risks undermining meaningful scrutiny and accountability, a significant concern when algorithmic systems are used in high-stakes decision-making.[3] If neither affected individuals nor regulators can understand how an AI reached a result, it becomes difficult to challenge or appeal unjust outcomes. Societies risk becoming dependent on these systems without truly understanding or regulating them, and especially in high-risk settings, this raises serious concerns about fairness, discrimination, and the use of sensitive data. For example, the algorithm used for criminal justice in the U.S. to predict a defendant's risk of reoffending, COMPAS, was criticised for its opaque logic and lack of accountability mechanisms. A 2016 investigation by *ProPublica* revealed that the COMPAS algorithm disproportionately classified Black defendants as high-risk compared to white defendants with more serious criminal records.[4] Accountability gaps resulting from opaque AI not only impair legal oversight but also threaten fundamental rights. When decisions affecting employment, liberty, or access to resources are made by inscrutable algorithms, principles of due process and fairness are put at risk. Thus, improving transparency and explainability in AI systems is widely seen as essential for enabling oversight and building public trust in AI-driven decisions.

1.3 Necessity for Regulation and the Post-GDPR Landscape

Given AI's ubiquity and the high stakes of its failures, there is a broad consensus that robust regulatory frameworks are needed to harness AI's benefits while mitigating its risks. The European Union's General Data Protection Regulation (GDPR) of 2018 marked a pivotal step in regulating automated data processing, setting strict standards for privacy, transparency, and individual rights. GDPR provides a solid foundation for AI governance, enshrining principles like data minimisation

[3] Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI' (2021) 41 Computer Law & Security Review 105567, 2–3.
[4]Duncan Purves and Jeremy Davis, 'Should Algorithms That Predict Recidivism Have Access to Race?' (2023) 60(2) *American Philosophical Quarterly* 131, 131–133.

and lawful processing that are directly relevant to AI systems.[5] It even anticipates algorithmic harms to an extent, for example, GDPR Article 22 establishes a right not to be subject to purely automated decisions without appropriate safeguards. However, GDPR was not designed as a comprehensive AI rulebook, and important aspects of AI fall outside its scope, such as governance of algorithmic fairness or safety in contexts beyond personal data.[6] As AI technologies evolve rapidly, policymakers have recognised that data protection law alone cannot address all emerging concerns. This has prompted new regulatory initiatives (such as the EU AI Act) aimed at a risk-based approach to AI, imposing obligations (like transparency, risk assessments, and human oversight) commensurate with an AI system's potential impact. Crucially, these initiatives strive to balance innovation with accountability. European experience shows that high standards for privacy and accountability need not stifle innovation: if viewed as an opportunity, measures like transparency and privacy-by-design can spur compliance-centred innovation that ultimately enhances user trust in AI products. Indeed, a policy study in the aftermath of GDPR's implementation concluded that effective AI governance should incentivise innovation that aligns with legal compliance, empower civil society to participate in AI oversight, and promote interoperability of AI governance across jurisdictions.[7] These lessons from the GDPR era underscore that thoughtful regulation, far from hindering technological progress, is necessary to ensure AI develops in a manner consistent with fundamental rights and societal values.

1.4 The Importance of Human Oversight

Across legal and ethical frameworks, human oversight has emerged as a cornerstone for trustworthy AI. However autonomous or "intelligent" an algorithm may be, human judgment and intervention

[5] Gianclaudio Malgieri, *Vulnerability and Data Protection Law* (Oxford University Press 2023) 165–175.
[6] Ibid 205–215.
[7] M Kuziemski and P Pałka, *AI Governance Post-GDPR – Lessons Learned and the Road Ahead* (European University Institute 2019) 27–30 https://data.europa.eu/doi/10.2870/470055

remain critical as a safety valve. Human oversight can take various forms, such as a human review of AI-generated decisions, real-time monitoring of an AI system's operation, or the ability to disable or override an AI in emergencies. The rationale is that meaningful human control is needed to prevent or correct AI errors and to inject ethical considerations that machines alone might overlook. This principle is now being codified in forthcoming regulations. Notably, under the EU AI Act, any system classified as high-risk AI must be designed and deployed with mechanisms for human oversight. Providers of such AI are obligated to ensure that humans can understand the system's functioning appropriately and intervene or stop the AI's operation if it malfunctions or produces harmful outcomes. Even GDPR requires that individuals subjected to automated decisions have the right to obtain human intervention and to contest those decisions. By mandating a human-in-the-loop, the law acknowledges that automated systems must ultimately remain subordinate to human agency and legal norms. In practice, effective oversight means not just having humans formally "in charge," but empowering them with the knowledge, training, and authority to evaluate AI outputs critically. Ensuring such oversight is vital to maintain accountability: it provides a check against opaque algorithms, helps uphold ethical standards, and reassures the public that AI is augmenting rather than replacing human decision-making in areas touching on rights and interests.

In sum, the ubiquity of AI in contemporary society and its opacity in decision-making processes have created urgent calls for governance that can safeguard public interests without unduly hampering innovation. Regulation is not a mere bureaucratic hurdle but a necessary instrument to steer AI development towards transparency, fairness, and respect for fundamental rights. Experience with data protection regimes like GDPR offers both a template and a cautionary tale: it provides principles to build on yet highlights the need for AI-specific rules and vigilant enforcement in the face of fast-evolving technologies. Central to this governance effort is retaining human oversight, ensuring that as we delegate tasks to machines, we do not abdicate accountability or moral responsibility. The following chapters will delve deeper into how legal frameworks, including data

protection law and emerging AI regulations, seek to address these challenges, aiming to chart a path where AI can be innovative, ubiquitous, and safe while operating under the rule of law and human conscience.

1.5 U.S. Approach

The AAA proposes to institutionalise impact assessments and stakeholder engagement in the United States, yet its legislative fate is uncertain. Against this backdrop, the thesis asks whether these instruments, individually and in concert with the AI Act, offer a satisfactory legal and ethical response to the challenges of AI-driven decision-making.

This research problem must also be situated in the context of contemporary policy debates. In mid-2025, the White House released America's AI Action Plan, a document framing AI as a global race and calling for the United States to innovate faster than its competitors. Rather than introducing new regulatory safeguards, the Plan champions a deregulatory agenda: it urges the dismantling of "**unnecessary regulatory barriers**". It warns that "**onerous regulation**" would not only unfairly benefit incumbents but could paralyse one of the most promising technologies. It further emphasises that AI systems should pursue "**objective truth**" and be free from ideological bias, and it instructs federal agencies to revise risk-management frameworks to eliminate references to misinformation, diversity, equity and inclusion, and climate change. By incorporating these developments into its analysis, the thesis highlights the ideological divergence between transatlantic approaches to AI governance, one emphasising risk management and fundamental rights and the other prioritising innovation, national security and deregulation and assesses how this divergence shapes the adequacy of existing legal frameworks.

2. Historical Context and Evolution

2.1 Evolution of EU Data Protection Law and the GDPR

2.1.1 Directive 95/46/EC – The Foundation of Modern Data Protection

The European Union's journey in data protection began with Directive 95/46/EC, commonly known as the Data Protection Directive. Adopted in 1995 after four years of negotiation, Directive 95/46/EC was the first comprehensive framework harmonising national data protection laws across EU Member States.[8] It had a dual objective: (1) to require every Member State to protect the fundamental rights and freedoms of individuals, in particular the right to privacy, with respect to personal data processing, and (2) to ensure the free flow of personal data within the EU single market without undue restrictions. These two goals were interlinked, seeking to guarantee a high level of personal data protection uniformly while facilitating the internal market's development.[9] The Directive drew heavily on the basic principles set out in the Council of Europe's Convention 108 of 1981, such as fairness, purpose limitation, and security, refining and supplementing them with further requirements. At the same time, because the Directive's provisions were broadly framed and left room for national implementation choices, Member States' laws, while more aligned than before, did not become identical.[10] Importantly, the 1995 Directive also broke new ground by introducing a rule on **automated individual decision-making** (Article 15), reflecting early recognition of risks posed by algorithmic decisions. Article 15 established that individuals should **not be subject to decisions significantly affecting them that are based solely on automated data processing,** subject to certain exceptions (such as decisions necessary for a

[8] Peter Hustinx, 'EU Data Protection Law: The Review of Directive 95/46/EC and the General Data Protection Regulation' in Paul Craig and Gráinne de Búrca (eds), *New Technologies and EU Law* (OUP 2017) 126, 126–127 https://doi.org/10.1093/acprof:oso/9780198807216.003.0005 accessed June 2025.
[9] Ibid 127-128.
[10] Ibid 128-129.

contract or authorised by law). This provision was motivated by the concern that computer-generated results might appear infallibly objective, leading human decision-makers to over-rely on them and abdicate responsibility[11]. By mandating a human review or intervention in such cases, the Directive set an early precedent for **algorithmic accountability** in European law.

Throughout the late 1990s and 2000s, Directive 95/46/EC served as the cornerstone of EU data protection. It influenced complementary instruments, for example, the e-Privacy Directive 2002/58/EC governing confidentiality in electronic communications and guided national laws and court decisions. However, rapid technological change (the rise of the internet, big data analytics, and eventually AI) exposed the Directive's limitations. Divergent national implementations led to fragmentation, and enforcement proved challenging. By 2012, the European Commission determined that an update was needed to strengthen rights and unify rules across the EU[12]. This led to the proposal and negotiation of the General Data Protection Regulation (GDPR), which would replace the 1995 Directive.

2.1.2 Data Protection as a Fundamental Right in the EU

A crucial development in this period was the formal recognition of data protection as a fundamental right. The Charter of Fundamental Rights of the EU, proclaimed in 2000 and given binding legal force in 2009 by the Lisbon Treaty, enshrines in Article 8 the right to the protection of personal data as a fundamental right distinct from privacy (which is protected separately in Article 7). The inclusion of Article 8 reflected the evolving European consensus that informational privacy deserved its own explicit guarantee. Indeed, with the Charter's entry into force, key principles of Directive 95/46/EC were effectively elevated to the level of primary law. [13]

[11] Lee A Bygrave, 'Automated Profiling: Minding the Machine – Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) 17(1) *Computer Law & Security Review* 17, 17–18.
[12] Hustinx, 'EU Data Protection Law (n8) 129-132.
[13] Ibid 138-140.

Article 8 of the Charter requires that personal data be processed fairly for specified purposes and on a legitimate basis laid down by law, and that everyone 1has the right of access to their data and the right to have it rectified. It also mandates independent supervision by an authority. These principles mirror core elements of the 1995 Directive (lawfulness, purpose limitation, data subject rights, and independent oversight), meaning that by 2009 the substance of EU data protection law had a constitutional status[14]. In parallel, the Treaty of Lisbon introduced Article 16 of the Treaty on the Functioning of the EU (TFEU), which explicitly authorises the EU to legislate on personal data protection for individuals (Article 16(1) TFEU) and provides a legal basis for adopting data protection rules through the ordinary legislative procedure (Article 16(2) TFEU)[15]. These changes cemented data protection as a fundamental value of EU law, influencing the ambition and scope of the GDPR.

The European Court of Justice (ECJ) quickly leveraged the Charter to strengthen data protection. In landmark rulings, the Court struck down EU laws that clashed with fundamental rights to privacy and data protection. For example, in Digital Rights Ireland (2014)[16], the ECJ invalidated the Data Retention Directive, finding that blanket retention of communications data disproportionately interfered with Articles 7 and 8 of the Charter. This jurisprudence underscored that any limitations on data protection must be necessary and proportionate, and it set the stage for the rigorous approach of the GDPR. The fundamental-rights perspective also meant that personal data transfers to third countries came under scrutiny through the Charter's lens (as discussed further below), ensuring that EU citizens' data would remain protected even when exported abroad. [17]

[14] Ibid 138-139.
[15] Ibid 141.
[16] Case C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others* [2014] ECLI:EU:C:2014:238
[17] Case C-362/14 *Maximillian Schrems v Data Protection Commissioner* [2015] ECLI:EU:C:2015:650.

By the early 2010s, it had become clear that the 1995 Directive's regime, crafted in an era when the internet was still nascent, was no longer adequate for the rapidly evolving digital society. Over two decades, the internet and algorithmic processing have transformed how data is collected and used, challenging the Directive's effectiveness. The patchwork implementation of the Directive by Member States also led to divergent rules and enforcement. Recognising these challenges, the European Commission launched a reform process in 2012 to **modernise and unify** EU data protection law. The result was the GDPR, adopted in 2016 and fully applicable from May 2018, replacing the 1995 Directive. Unlike a directive, the GDPR is a regulation that is directly applicable in all Member States, thereby eliminating national discrepancies. EU institutions heralded the GDPR as one of the Union's great achievements, a comprehensive update fit for the big data era.[18] The GDPR not only carried forward the objectives of its predecessor but also significantly **strengthened individual rights and enforcement mechanisms**, as discussed below.

2.1.3 Foundational Principles and EU Values Underpinning the GDPR

The GDPR was born out of a deliberate effort to reinforce European **fundamental values** in the digital age. Foremost among these is the respect for fundamental rights, including **the right to privacy** and **the protection of personal dat**a as guaranteed by the EU Charter. The very first recital of the GDPR affirms that "*the protection of natural persons in relation to the processing of personal data is a fundamental right,*" echoing Article 8 of the Charter.[19]

This central premise demonstrates continuity with Europe's longstanding view of privacy and data protection as core aspects of human rights rather than mere consumer or economic interests. Indeed, the European approach to data protection is deeply linked to the concept of human dignity. In the

[18] European Data Protection Supervisor, *The History of the General Data Protection Regulation* https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en accessed August 2025.
[19] Regulation (EU) 2016/679 … (General Data Protection Regulation) [2016] OJ L119/1, recital 1.

aftermath of World War II, and informed by the horrors of totalitarian misuse of personal information, European legal thought tied privacy closely to the protection of human dignity.[20] This is exemplified in constitutional traditions (notably Germany's jurisprudence on "informational self-determination") and international texts. The EU Charter's Article 1 places human dignity as inviolable and foundational, and privacy and data protection are seen as instrumental to preserving individual autonomy and personhood. Thus, the driving principles behind the GDPR's enactment were not only pragmatic or economic, but fundamentally **ethical and rights-oriented**, seeking to preserve individual control over personal information in an age of massive data flows.

Crucially, the GDPR's legislative history and content reflect an intent to **strengthen individuals' control and agency** vis-à-vis those who process their data. During the reform process, EU policymakers emphasised rebuilding trust in the digital economy by enhancing transparency and accountability in data processing. Although a detailed analysis of principles such as **lawfulness, transparency, purpose limitation, data minimisation, fairness, and accountability** is reserved for Chapter 3, it is worth noting that these principles underlie the GDPR's provisions. For example, the GDPR introduced or bolstered data subject rights (such as t**he right to access, rectification, erasure, and data portability**) and placed stricter conditions on consent, all aimed at empowering individuals. It also considerably expanded compliance obligations for organisations (e.g. the principles of "*privacy by design/default"* and requirements for data protection impact assessments) and provided for **robust enforcement** (with independent supervisory authorities and the threat of heavy fines for violations). These changes were driven by the understanding that personal data has become central to individuals' lives and liberties, and thus, strong safeguards are needed to uphold the EU's fundamental values in practice.[21]

[20] Sylvia Lissens, *The Foundations of EU Personal Data Protection Law: Privacy and Human Dignity* (EU-RENEW, 30 January 2024) https://eu-renew.eu/the-foundations-of-eu-personal-data-protection-law-privacy-and-human-dignity/
[21] European Union Agency for Fundamental Rights, *GDPR in practice – Experiences of data protection authorities* (11 June 2024) https://fra.europa.eu/en/publication/2024/gdpr-experiences-data-protection-authorities

Notably, the reform coincided with the EU's broader recognition of data protection as integral to human dignity and democracy: European law treats certain intrusive data practices and automated profiling as "an inherent threat to human dignity," an approach quite distinct from more laissez-faire traditions elsewhere.[22] In summary, the GDPR's enactment was guided by the dual imperative of **protecting individual rights** in the information age while **enabling the free flow of data for the digital single market,** with a balance rooted in human dignity, autonomy and solidarity.

2.2 Automated Decision-Making and Data Transfers

Two areas of legal development that are particularly relevant to AI-powered decision-making systems have been: **(a)** protections against harmful automated decisions, and **(b)** rules on international data transfers.

2.2.1 Automated decision-making

Notably, the 1995 Directive itself contained a forward-looking provision (Article 15) granting individuals the right not to be subject to a decision with significant effects based solely on automated processing. This little-invoked article, inspired in part by early member-state laws (France's 1978 data protection law famously prohibited purely automated administrative decisions by default), reflected a European wariness that removing human judgment from important decisions could threaten individual dignity and fairness.[23] The GDPR built upon this foundation: it introduced **Article 22,** which provides that individuals "**have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning them or similarly significantly affects them**", unless a specified exception applies (such as

[22] Bygrave (n 11), 18-19.
[23] Bygrave (n 11), 17-19.

explicit consent, necessity for a contract, or authorisation by law with suitable safeguards).[24] GDPR Article 22 thus carries forward the spirit of its predecessor but with clearer scope and stronger safeguards. It establishes a default prohibition on purely automated decisions with significant impact, reflecting the European view that such AI-driven decisions are inherently suspect and should be allowed only under strict conditions. In situations where automated decision-making is permitted, the GDPR requires that data subjects receive meaningful information about the logic involved and that they have the right to obtain human intervention, express their point of view, and contest the decision (Articles 13–15 and Article 22(3) GDPR). These provisions are essentially aimed at ensuring **human accountability and transparency** in algorithmic decisions.

Despite these legal protections, debate has continued about their sufficiency in the age of advanced AI. One much-discussed issue is whether the GDPR implicitly guarantees a "right to explanation" of automated decisions. While the GDPR stops short of explicitly mandating a full explanation, Recital 71 and supervisory guidance have emphasised that any automated decision-making must be transparent and fair, and individuals should be able to seek an explanation of the decision-making logic as part of exercising their rights. The question of how far these rights extend is now being tested in practice. A notable recent development is the case of **SCHUFA credit scoring** in Germany, which has brought automated decision-making before the CJEU for the first time. In that case **(C-634/21, OQ v. Land Hessen)**, the issue is whether a credit bureau's purely algorithmic score, used by lenders to approve or deny loans, counts as an automated decision producing legal or similarly significant effects under GDPR Article 22(1). In March 2023, Advocate General Pikamäe advised that the act of generating and disseminating a credit score should indeed be treated as an automated decision within the meaning of Article 22, given its crucial influence on third-party decisions like loan approvals.[25] If the CJEU follows this opinion, it will affirm a broad

[24] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1, art 22.
[25] Boris Paal, 'Article 22 GDPR: Credit Scoring Before the CJEU' (2023) 4(3) *Global Privacy Law Review* 127, 127-133.

interpretation of individuals' rights in the context of AI-driven profiling, reinforcing that even intermediary algorithmic outputs (not just final decisions) can trigger GDPR protections. This pending decision represents a key moment in clarifying and evolving the legal framework for AI accountability under data protection law.

2.2.2 Data transfers

The cross-border flow of personal data, especially to jurisdictions without equivalent privacy laws, has been a persistent concern and has seen dramatic legal developments over the past decade. Under both Directive 95/46/EC and the GDPR, transfers of personal data from the EU to third countries are generally permitted only if the destination country ensures an "**adequate**" level of protection or if appropriate safeguards (such as standard contractual clauses or binding corporate rules) are in place.[26]

A major milestone was the European Commission's **Safe Harbour decision (2000)**, which found that U.S. companies complying with a set of self-regulatory principles would be deemed to provide adequate protection. Safe Harbour facilitated EU–U.S. data flows for years until it was invalidated by the CJEU in the 2015 **Schrems I** ruling. In Schrems I, an Austrian individual (Max Schrems) challenged Facebook's transfers of EU user data to the U.S., arguing that U.S. surveillance practices undermined the promised protections. The CJEU agreed: on 6 October 2015, it struck down the Safe Harbour framework, holding that the Commission's adequacy decision was invalid because it did not ensure essentially equivalent protection for EU citizens' data in the U.S.[27] The Court emphasised that EU data exporters and national data authorities must stop transfers if foreign law or

[26] Sergi Batlle and Arnaud van Waeyenberge, 'EU–US Data Privacy Framework: A First Legal Assessment' (2024) 15(1) *European Journal of Risk Regulation* 191, 192.
[27] Shara Monteleone and Laura Puccio, *The CJEU's Schrems Ruling on the Safe Harbour Decision* (European Parliamentary Research Service, October 2015) https://www.europarl.europa.eu/RegData/etudes/ATAG/2015/569050/EPRS_ATA(2015)569050_EN.pdf

practices (such as indiscriminate government access to data) compromise fundamental privacy rights. [28] This judgment underscored that high EU data protection standards "travel with the data" and cannot be disregarded overseas.

In response to Schrems I, the EU and U.S. negotiated a replacement arrangement, the **Privacy Shield (2016)**, which introduced additional safeguards and U.S. government assurances. However, in July 2020, the CJEU issued the **Schrems II** decision (Case C-311/18) and again invalidated the EU's adequacy decision for Privacy Shield. The Court found that U.S. surveillance laws (notably Section 702 FISA and Executive Order 12333) still allowed intelligence agencies to collect Europeans' data on a broad scale without sufficient judicial oversight or redress for EU individuals. [29] It concluded that the U.S. framework did not meet the "essential equivalence" standard required by Article 45 GDPR, particularly because affected individuals lacked an effective remedy before an independent tribunal. Schrems II sent shockwaves through the industry and forced companies to rely on standard contractual clauses and other transfer mechanisms, while simultaneously pressuring the EU and U.S. to craft a new solution. In 2023, the European Commission approved the **EU–U.S. Data Privacy Framework**, a revamped adequacy decision aiming to address the CJEU's concerns (for example, by introducing a new redress mechanism via a Data Protection Review Court in the U.S.). Whether this new framework will survive legal challenge remains to be seen, but these episodes have been pivotal in the evolution of EU data protection. They highlight the primacy of fundamental rights in EU digital policy and have spurred both stricter data transfer compliance and a greater push for technological sovereignty in Europe. Crucially, the Schrems rulings also carry implications for AI governance: many AI systems rely on large datasets and cloud infrastructure that span global borders, so the EU's insistence on fundamental-rights adequacy in data transfers is an integral part of ensuring that AI-driven decisions involving personal data are subject to EU standards, no matter where processing occurs.[30]

[28] Ibid.
[29] Batlle and van Waeyenberge (n 26), 193.
[30] Ibid.

In summary, by the time the GDPR took effect in May 2018, it stood on firm historical foundations: decades of data protection principles from Directive 95/46/EC, newly constitutionalised by the EU Charter, and reinforced by case law defending privacy and personal data as fundamental rights. These developments established a robust baseline for governing algorithmic decision-making. The GDPR's provisions on transparency, data subject rights, and automated decisions provide important tools that continue to apply to AI-powered systems today. However, as the next section details, the EU determined that the GDPR alone was not sufficient to address all the challenges of AI, particularly those not strictly related to personal data, leading to the proposal of a dedicated Artificial Intelligence Act.

2.3 Development of the EU AI Act

Legislative Debates

In the mid-2010s, as artificial intelligence technologies advanced rapidly, EU institutions began exploring a regulatory approach specific to AI. While the GDPR addresses certain concerns (notably privacy and some automated decisions), its scope is limited to personal data processing. Broader issues posed by AI systems, such as safety, transparency, bias, and societal risks, even where no personal data is involved, prompted calls for bespoke legislation. The **European Commission's White Paper on AI (February 2020)** set out policy options, and an expert group (the High-Level Expert Group on AI) had earlier proposed ethical guidelines. Building on this groundwork, the Commission in April 2021 published its proposal for a comprehensive AI-specific regulation, which would eventually become the **Artificial Intelligence Act (AI Act)**. The proposal immediately sparked extensive political debate, both within the EU's legislative bodies and among stakeholders across industry, civil society, and academia.

A central question in negotiations was how to achieve an appropriate balance between **mitigating AI's risks** and **not stifling innovation**. The Commission's draft embraced a **risk-based regulatory model**, as can be seen in Figure 1[31], categorising AI systems by the level of risk they pose (unacceptable risk, high-risk, limited risk, minimal risk) and imposing obligations accordingly.

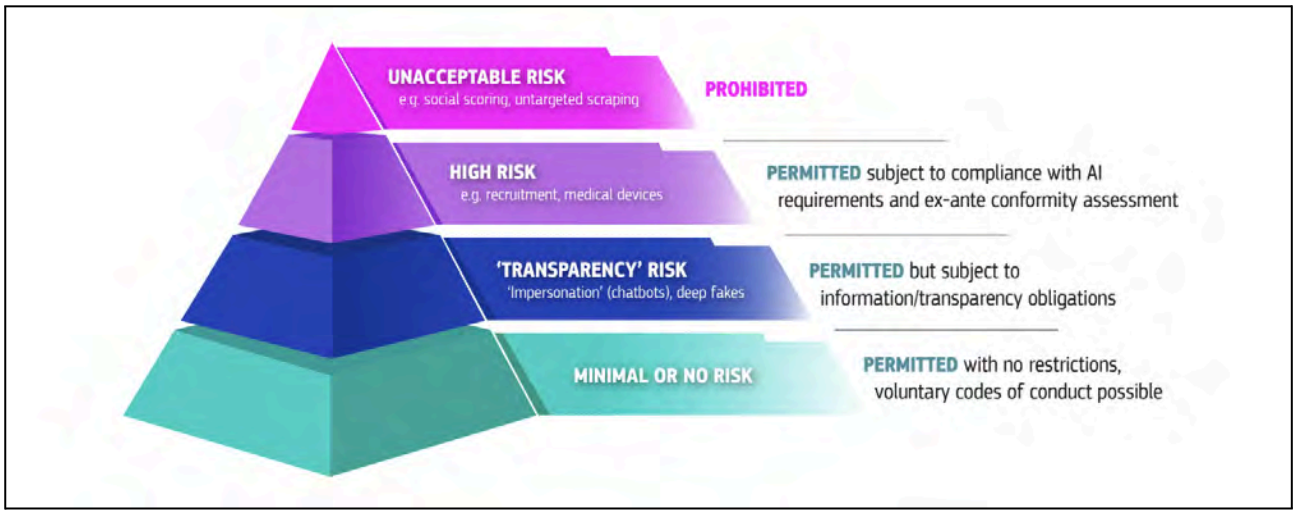

Figure 1: EU AI Act Risk-Based Pyramid

This structure drew broad support, but there were intense debates on the details. One contentious area was the scope of **prohibited AI practices**. The initial proposal banned a set of AI uses seen as contravening EU values (such as social scoring by governments, exploitative use of vulnerabilities, or real-time remote biometric identification in public by law enforcement). The European Parliament, influenced by human rights concerns, pushed to expand the list of bans; for example, many MEPs advocated a stricter prohibition on facial recognition surveillance in public spaces. On the other hand, some Member States in the Council were wary of outright bans that might hinder law enforcement or security uses. The **political compromise** (reached in late 2023) maintained a ban on social scoring and indiscriminate biometric surveillance, but it **allows certain uses of facial recognition by police with judicial authorisation and safeguards**.[32] This was seen as a middle

[31] European Commission, *AI Act* (EU Digital Strategy, 30 July 2024), image 'AI pyramid of risks according to AI Act'
[32] Council of the EU, Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World(Press Release 986/23, 9 December 2023, updated 2 February 2024)

ground between Parliament's more restrictive stance and the Council's preferences, reflecting broader debates on how far AI regulation should go in limiting government use of AI tools.

Another major debate concerned **"general-purpose AI" (GPAI) and foundation models,** AI systems like large language models that can be adapted to countless tasks (e.g. GPT-type models). These technologies exploded in capability and public awareness during the legislative process (notably with ChatGPT's debut in 2022), catching regulators' attention. The European Parliament pressed for the AI Act to explicitly cover foundation models and generative AI, introducing provisions during 2023 that would require makers of such models to implement extra transparency and risk management steps. The final Act, as agreed in 2024, includes specific obligations for providers of foundation models and generative AI (for instance, requiring disclosure that AI-generated content is AI-made, and mandating risk assessment for models that could pose systemic societal risks). [33] These provisions were politically significant: they marked the world's first attempt to directly regulate the developers of large AI models, not just the downstream deployers. Reaching an agreement required addressing industry concerns about feasibility and international competitiveness, and the Act ultimately built in grace periods and soft-law tools (like a voluntary Code of Conduct for GPAI providers) to smooth implementation. [34]

The **governance and enforcement structure** was another point of negotiation. The Commission's proposal largely relied on Member State authorities (mirroring how GDPR is enforced by national data protection authorities). However, given the technical complexity of AI, the Parliament favoured a stronger central coordination role. The compromise establishes a hybrid system: each Member State will designate national AI supervisors, but there will also be a new **European AI Board/Office** to facilitate consistent application and to wield certain oversight powers at the EU level. This echoes the GDPR's mechanism of a European Data Protection Board, while granting the

[33] European Commission, 'AI Act' (*Shaping Europe's Digital Future*) (last updated 1 August 2025) https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[34] Ibid.

AI Board a mandate to issue guidance and coordinate enforcement for AI systems that have a cross-border impact.

By December 2023, after lengthy trilogues, EU lawmakers announced a provisional political agreement on the AI Act hailed as "*the world's first comprehensive AI regulation*". This deal was finalised in 2024. The Council's press release at the time called it a "historical achievement" after a "3-day marathon" negotiation, highlighting that the Act aims to address a global challenge by ensuring AI in Europe is "safe and respects fundamental rights and EU values".[35] Notably, policymakers were keen to draw parallels to the GDPR's global influence, suggesting the AI Act could similarly become a de facto international standard.[36] The Act was formally approved by the European Parliament and Council in mid-2024 (Regulation (EU) 2024/1689), and it entered into force on 1 August 2024. [37] However, recognising the adaptation needed, its provisions have staggered application dates: the bulk of obligations (especially for high-risk systems) will apply after a two-year implementation period, i.e. starting mid-2026, with some measures on prohibited practices and general-purpose AI kicking in earlier.[38]

2.4 The United States' Legislative Approach to AI

2.4.1 Overview of U.S. Federal AI Policy Developments

In the absence of a comprehensive federal law on artificial intelligence (AI), the United States has approached AI governance through a patchwork of policy initiatives, executive actions, and proposed legislation. Early federal attention emphasised AI innovation and competitiveness over regulation. During the Trump administration's first term (2017–2021), the White House launched

[35] Council of the EU (n 31)
[36] Ibid.
[37] European Commission, 'AI Act' (*Shaping Europe's Digital Future*) (n 32)
[38] Ibid.

the **American AI Initiative** via executive order in 2019, aiming to boost AI R&D and preserve U.S. leadership, while cautioning regulators against stifling innovation with overbroad rules.[39] This innovation-centric approach largely avoided new legal constraints on AI, reflecting a view that existing laws and market forces would suffice to address issues like bias or privacy. By contrast, as AI technologies (especially machine learning) began demonstrating pervasive societal impacts, calls grew for a more proactive regulatory stance grounded in rights and accountability.

Congressional activity on AI picked up markedly in the late 2010s and early 2020s. Lawmakers formed AI-focused caucuses and committees convened expert hearings, and introduced numerous bills to study or guide AI development. [40] Yet no broad AI-specific statute has been enacted to date. Instead, Congress incorporated modest AI provisions into other laws. For example, the **National AI Initiative Act of 2020** established coordination bodies and research programs but imposed no new private sector mandates. During the 118th Congress (2023–2024), Senate leadership convened a Bipartisan AI Working Group that held high-profile "AI Insight Forums," producing a 2024 roadmap with policy recommendations on AI risks, transparency, workforce impacts, and safeguards. In parallel, a House bipartisan Task Force on AI studied potential "guardrails" to balance innovation with protection of rights and safety. These efforts signalled a growing bipartisan consensus that federal intervention may be needed to address AI's societal implications.[41]

At the executive branch level, the Biden administration (2021–2024) gradually pivoted federal AI policy toward issues of **trust, safety, and civil rights**, especially after the public debut of advanced systems like ChatGPT in 2022.[42] Notably, the White House Office of Science and Technology

[39] Executive Order No 13859, *Maintaining American Leadership in Artificial Intelligence* (11 February 2019) 84 Fed Reg 3967; Office of Management and Budget, *Guidance for Regulation of Artificial Intelligence Applications* Memorandum M-21-6 (17 November 2020).
[40] Laurie Harris, *Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress*(Congressional Research Service, 4 June 2025) https://www.congress.gov/crs_external_products/R/PDF/R48555/R48555.2.pdf
[41] Ibid.
[42] Sarah Box, *The Road Ahead for U.S. Policy on Artificial Intelligence* (Background Paper No 31, ORF America, 31 March 2025) 7-8 https://orfamerica.org/newresearch/us-policy-on-artificial-intelligence

Policy in October 2022 released a Blueprint for an **AI Bill of Rights,** outlining principles for safe and effective systems, algorithmic nondiscrimination, data privacy, notice and explanation, and human alternatives. This Blueprint, while not legally binding, reflected a human-rights-centric vision for AI governance. It was followed in October 2023 by an unprecedented Executive Order 14110 on "**Safe, Secure, and Trustworthy Development and Use of AI**," which declared that AI "holds extraordinary potential for both promise and peril" and set forth a sweeping federal agenda for "responsible AI."[43]

The order directed over 50 agencies to take around 150 actions, aiming to foster AI safety standards,

address algorithmic bias, protect privacy, and ensure human oversight in critical uses of AI. For instance, it instructed the Commerce Department to require developers to report when training advanced AI models that pose serious risks, and urged agencies to use existing consumer protection and civil rights laws to crack down on AI-driven fraud or discrimination. It also announced the formation of an AI Safety Institute within NIST to develop risk mitigation guidelines. However, without new legislation, these measures relied on agencies' existing powers or voluntary cooperation from industry, and the rights articulated (e.g. the right to notice or to opt out of automated decisions) remained aspirational rather than enforceable by the public.[44]

By early 2025, the contrast between regulatory philosophies became even starker with a change in administration. On his first day in office (January 20, 2025), President Donald Trump (in his second term) revoked the 2023 Biden AI order, which his party's platform had decried as "dangerous."[45] Three days later, he issued a new Executive Order entitled "**Removing Barriers to American**

---

[43] Executive Order No 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (30 October 2023) 88 Fed Reg 73589.
[44] Adam Aft, Caroline Burnett, Cynthia Cole, Brian Hengesbaugh, Keo McKenzie, Justine Phillips, Elizabeth Roper and Avi Toltzis, 'AI Tug-of-War: Trump Pulls Back Biden's AI Plans' (The Employer Report, 25 January 2025) https://www.theemployerreport.com/2025/01/ai-tug-of-war-trump-pulls-back-bidens-ai-plans/
[45] Ibid.

Leadership in AI," refocusing U.S. AI policy on deregulation and dominance. Trump's order set a tone of strategic competition, declaring it national policy to "sustain and enhance America's global AI dominance" in the service of economic competitiveness and security. It rescinded any pending agency actions from Biden's order deemed to "hamper U.S. leadership," and it pointedly omitted Biden's emphasis on civil rights or consumer protections.[46] Instead, the Trump administration convened an interagency process to develop "**Winning the AI Race: America's AI Action Plan**," released in July 2025, which doubled down on AI innovation through measures like expanding R&D infrastructure, easing regulatory burdens, and promoting AI exports[47]. The Action Plan's pillars, *Accelerating Innovation, Building AI Infrastructure, and International Leadership*, underscore a clear priority on outpacing geopolitical rivals in AI. While it included references to avoiding "Orwellian" uses of AI and centring American workers, its concrete policies (e.g. fast-tracking data centre permits, requiring federal AI procurements to favour "unbiased" models) primarily furthered a pro-business agenda.[48] In short, the absence of a statutory framework left U.S. AI governance swinging with the political pendulum: one administration's voluntary guidelines for ethical AI were easily undone by the next, in favour of laissez-faire approaches. This volatility has intensified calls for Congress to establish a durable, rights-based federal law on AI that transcends executive preferences.

2.4.2 The Algorithmic Accountability Act: Evolution and Key Provisions (2019–2025)

In the United States, awareness of biases and harms in AI-driven decision-making has grown in recent years, but comprehensive federal legislation has lagged behind the EU. One high-profile attempt to fill this gap was the **Algorithmic Accountability Act (AAA)**, originally introduced in 2019 at the federal level. In April 2019, Senator Ron Wyden, with co-sponsors including Senator

[46] Ibid.
[47] *America's AI Action Plan* (The White House, July 2025) 3-5 https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf
[48] Ibid.

Cory Booker, and Rep. Yvette Clarke in the House, unveiled the Algorithmic Accountability Act of 2019. Prompted by rising concern over "secret" algorithms making decisions in housing, employment, credit and other life opportunities, the 2019 bill sought to "pull back the curtain" on automated decision systems that could produce discriminatory or unfair outcomes.[49]

This bill sought to empower the Federal Trade Commission (FTC) to require companies to conduct algorithmic impact assessments on their automated decision systems, particularly those impacting consumers' rights and opportunities (such as credit, employment, insurance, or housing decisions). The 2019 bill defined automated decision systems broadly and would have applied to companies above certain size thresholds or engaging in large-scale data practices.[50] The required impact assessments were meant to evaluate the design and outcomes of AI systems for **accuracy, fairness, bias, privacy, and security issues, and to mitigate any identified risks**. In essence, the Act attempted to introduce accountability for AI at the organisational process level: companies would need to audit their algorithms and document steps taken to avoid discriminatory or other harmful effects. The AAA of 2019 garnered significant public interest and it was among the first pieces of U.S. legislation to directly address algorithmic bias across sectors.

While the 2019 proposal did not advance out of committee, it established a template for algorithmic accountability, namely, mandating **internal due diligence** by companies and oversight by the FTC. The bill's introduction alone was a landmark; it put Congress on record recognising that algorithms can "**decide whether Americans get to see a doctor, rent a house or get into a school,**" and that industry's "blind eye" toward algorithmic bias was unacceptable. [51]

---

[49] Wyden, Booker and Clarke, 'Wyden, Booker and Clarke Introduce Algorithmic Accountability Act of 2022 To Require New Transparency and Accountability for Automated Decision Systems' (Press Release, US Senate, 3 February 2022) https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems

[50] Algorithmic Accountability Act of 2019, HR 2231, 116th Congress (2019)

[51] Wyden, Booker and Clarke (n 49).

Revisions in 2022 and 2023: Lawmakers reintroduced the AAA in the 117th Congress (as the Algorithmic Accountability Act of 2022) with some updates informed by intervening debates. The 2022 bill (S.3572 and H.R.6580) explicitly covered not only AI-driven decisions by the companies themselves but also algorithms sold or provided by one company for use in another entity's decision-making ("augmented critical decision processes").[52] This closed a loophole, ensuring that vendors of algorithmic tools would also need to assess and disclose their products' impacts. The 2022 version sharpened the focus on "**critical decisions**" defined as decisions affecting legal rights or important life opportunities in areas like housing, employment, education, healthcare, finance, or access to basic utilities.[53] By tailoring the law to these high-impact domains, the Act targeted algorithms most likely to implicate civil rights and fairness.

Key provisions of the AAA (2022–2023) included:

- **Mandatory Impact Assessments:** Covered entities must perform an **Algorithmic Impact Assessment (AIA)** for each automated decision system used in a critical decision process, both before deployment and periodically thereafter. The assessment requires evaluating the system's purpose, its training data, and its performance on metrics including **accuracy, fairness (bias and nondiscrimination), privacy and security, transparency and explainability**, and other factors. Importantly, firms have to examine whether the AI system produces or exacerbates biased outcomes for any protected or marginalised groups. They must also test the system's robustness and reliability (e.g. susceptibility to errors or manipulation), and consider whether affected individuals have an opportunity to contest or seek recourse for an automated decision. If any aspect of these assessments is not feasible, for example, if the company lacks the necessary information about a third-party model or cannot obtain relevant demographic data, the firm must

[52] Algorithmic Accountability Act of 2022, HR 6580, 117th Congress (2022)
[53] Harris (n 40).

document these gaps and why they could not be remedied. This creates a paper trail to prevent willful ignorance of an algorithm's impacts.

- **Documentation and Reporting:** The AAA requires companies to maintain documentation of each impact assessment for at least 3 years beyond the AI system's use. Critically, firms must submit summary reports of their assessments to the FTC. Under the 2022 bill, any new high-risk automated system triggers an initial report to the FTC prior to deployment, and ongoing systems require annual update reports. The FTC, in turn, is charged with creating a publicly accessible repository of these algorithmic systems, publishing a subset of information from the company reports. In essence, the law would generate a first-of-its-kind federal registry of algorithms used in sensitive domains, bringing transparency to what has largely been an opaque private sector practice. The FTC would also receive funding to hire at least 75 additional staff and even establish a new Bureau of Technology to assist with technical evaluations and enforcement. This acknowledgement of needed capacity is notable, as it recognises that regulators require data scientists and AI experts to effectively oversee compliance.
- **Enforcement Mechanism:** A violation of any FTC regulation under the AAA (for example, failing to perform an assessment or to file a required report) is deemed an "unfair or deceptive act or practice" under the FTC Act, enforceable by the Commission with civil penalties. The AAA thus leverages the FTC's existing enforcement powers, but bolsters them with explicit new duties and resources. However, the Act stops short of creating a private right of action individuals would not directly sue companies under the AAA, but rather would rely on the FTC to police industry behavior. The expectation is that the threat of FTC investigation and penalties, combined with public transparency, would incentivize companies to proactively address issues the assessments uncover.[54]

[54] Algorithmic Accountability Act of 2019, HR 2231, 116th Congress (2019); Algorithmic Accountability Act of 2022, HR 6580, 117th Congress (2022); Algorithmic Accountability Act of 2023, HR 5628, 118th Congress (2023)

Through these provisions, the AAA bills aimed to embed **accountability-by-design** into AI deployment. Senator Wyden summarized the goal as giving consumers "knowledge of when and how they've been impacted" by automated decisions and ensuring companies **"assess the impact of their algorithms and hold bad actors accountable"**.[55] By requiring documentation of how an algorithm works and affects people, the AAA aligns with the principle of **explicability** (the idea that AI outcomes should be intelligible and answerable to the public). And by mandating bias testing and data protection reviews, it seeks to safeguard **equality and privacy rights** in automated systems.

2025 Version: In the 119th Congress, lawmakers introduced an updated Algorithmic Accountability Act of 2025 (S. 2164, 119th Congress), continuing the effort. Largely, it carried forward the core structure of the 2022–23 bills. One noteworthy refinement in 2025 was an even more explicit inclusion of algorithmic supply chains: companies providing AI systems to others would have to disclose to their clients whether they qualify as a covered entity under the Act.[56] This provision is meant to alert downstream users that a given tool is subject to regulatory scrutiny and reporting. The 2025 bill also maintained the emphasis that simply outsourcing an automated decision to an external vendor does not exempt the deploying company from responsibility.

Despite bipartisan recognition of AI risks, the AAA bills (2019 through 2025) have not been enacted. Some opposition centres on concerns that mandatory assessments and reporting could impose compliance costs and slow innovation, especially for smaller firms, though the Act's thresholds are intended to exempt small businesses. For example, earlier drafts applied only to companies with over $50 million in revenue or dealing with personal data on >100k individuals, focusing on larger data-centric firms.[57] Industry lobbyists have also argued that requiring disclosure

[55] Wyden, Booker and Clarke (n 49).
[56] Algorithmic Accountability Act of 2025, S 2164, 119th Congress (2025)
[57] Algorithmic Accountability Act of 2023, Discussion Draft, 118th Congress (Sen Ron Wyden, 2023) § 2(7)(A) The AAA 2019 applied to companies meeting thresholds in the definition of 'covered entity', such as having over

of algorithmic details might risk revealing proprietary information or could be ineffectual if not paired with clear rules on outcomes. Indeed, a critical limitation of the AAA approach is that it does not prohibit any particular uses of AI or outcomes. As the Congressional Research Service observed, the AAA's framework would mandate evaluation and reporting of AI systems' impacts, "*and did not contain specific prohibitions on use*." [58] In other words, a company could theoretically conduct an impact assessment that reveals bias or privacy problems, file the report, and still go live with the system. The Act relies on sunlight and FTC oversight rather than direct bans or new civil liability. This contrasts with, for example, the EU's approach of outright forbidding certain high-risk AI practices. Nonetheless, the AAA represents a significant step toward accountability: it creates mechanisms to expose and address harm before it occurs (through pre-deployment assessment) and to hold organizations answerable for the algorithms they operate.

In sum, the AAA bills reflect an incremental but important evolution in U.S. legislative thinking. From 2019 to 2025, they expanded in scope and specificity, responding to the public's growing demand for **algorithmic transparency and fairness.** They also signalled a commitment to democratic oversight of AI, vesting a public agency (FTC) with insight into systems that were previously black boxes. Although not yet law, the AAA's concepts have influenced other proposals and state-level efforts, and they provide a blueprint for any future comprehensive AI regulation in the United States that seeks to balance innovation with fundamental rights.

2.4.3 State-Level Responses

In the absence of enacted federal legislation, a number of U.S. states and cities have stepped in to address AI and algorithmic decision-making within their jurisdictions. These efforts are creating a patchwork of algorithmic accountability rules aimed at specific sectors or issues. One pioneering

$50,000,000 in annual gross receipts or possessing personal information of over 100,000 consumers or devices. Later versions retained similar concepts, ensuring startups and small businesses were not overburdened.

[58] Harris (n 40).

example is **Colorado's Senate Bill 21-169 (SB21-169)**, enacted in July 2021, which represents one of the first state laws explicitly regulating the use of AI-driven tools for fairness. SB21-169, titled "**Restrict Insurers' Use of External Consumer Data and Information**," was designed to prevent unfair discrimination in the insurance industry's use of algorithmic models.[59] Under this law, insurers in Colorado are prohibited from using any "external consumer data and information source, algorithm, or predictive model" in insurance practices that result in unfair discrimination on the basis of protected characteristics such as race, colour, national origin, religion, sex, sexual orientation, disability, or gender identity. In essence, if an insurance company deploys a machine-learning model (for, say, setting rates or determining eligibility) that ends up treating people differently due to those sensitive attributes (even indirectly, via proxies), that would violate the law.

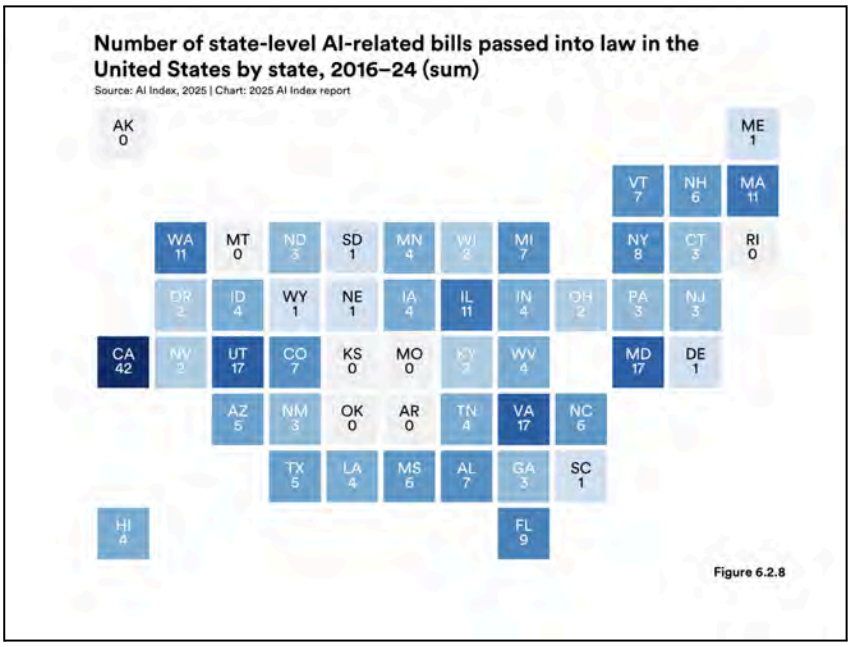

Figure 2: Number of state-level AI-related bills passed into law in the United States, 2016–24

As illustrated in Figure 2[60], the United States has witnessed a fragmented but accelerating development of AI-related legislation at the state level. California leads by a significant margin, with forty-two AI-related bills enacted between 2016 and 2024, followed by Utah, Virginia, and Maryland, each with seventeen. By contrast, many states, such as Kansas, Missouri, and Arkansas, have passed no AI-specific laws during the same period. This uneven distribution highlights the absence of a unified federal framework, leaving states to experiment with their own regulatory approaches. While this

[59] Colorado Senate Bill 21-169 (Restrict Insurers' Use of External Consumer Data and Information) (2021)
[60] AI Index, *2025 AI Index Report* (Stanford Institute for Human-Centered Artificial Intelligence 2025) fig 6.2.8.

demonstrates legislative dynamism, it also raises concerns about regulatory fragmentation, legal uncertainty for interstate businesses, and inconsistencies in protecting individuals' rights across jurisdictions.

This chapter has traced the evolution of AI governance on both sides of the Atlantic, highlighting the EU's structured regulatory development and the United States' fragmented approach, shaped by shifting executive priorities, congressional proposals such as the Algorithmic Accountability Act, and emerging state-level initiatives. This comparative context underlines two key insights for the present study: first, that the legal architecture for AI-powered decision-making remains incomplete in all jurisdictions; and second, that foundational legal and ethical principles are indispensable to bridge regulatory gaps and guide future legislative design. The following chapter, therefore, turns to a detailed examination of these principles anchored in binding instruments such as the GDPR and the EU AI Act, informed by U.S. proposals like the AAA, and enriched by international ethical frameworks to assess how they can collectively ensure that AI systems operate transparently, fairly, and in alignment with fundamental rights.

3. Legal and Ethical Principles

3.1 Core Principles of the GDPR

The GDPR is built on a set of fundamental data-protection principles intended to safeguard individuals' rights and regulate how personal data is used. In the context of AI-powered decision-making systems, three core principles are particularly prominent: **lawfulness of**

processing, **purpose limitation**, and **data minimisation**. These principles, secured in Articles 5 and 6 GDPR, define ex ante obligations for controllers deploying AI models that rely on personal data. This section examines each principle's doctrinal basis and legal obligations, then analyses their application to AI model training and deployment.

3.1.1 Lawfulness of Processing (Article 6 GDPR)

Article 5(1)(a) GDPR establishes that personal data must be processed "**lawfully, fairly and in a transparent manner**".[61] The lawfulness requirement is fleshed out in Article 6(1), which provides an exhaustive list of lawful bases for processing. Accordingly processing is lawful only if at least one of the following applies: **(a)** the data subject has given **consent** for specific purposes; **(b)** it is necessary for performance of a **contract** with the data subject; **(c)** it is necessary for compliance with a **legal obligation**; (d) it is necessary to protect someone's **vital interests**; (e) it is carried out in the **public interest** or under official authority; or (f) it is necessary for the **legitimate interests** of the controller or a third party, provided this is **not overridden by the individual's interests or fundamental rights**.[62] In AI development, where vast amounts of personal data may be ingested, identifying an appropriate legal basis is a critical first step.

Applying Lawful Bases to AI Systems: AI-powered systems like facial recognition or algorithmic credit scoring often involve processing personal data on a large scale, sometimes collected indirectly or repurposed from other contexts. In practice, obtaining explicit consent for every data subject whose data is used to train an AI model can be infeasible or impracticable, for example, scraping millions of images for a facial recognition dataset. Controllers therefore frequently rely on the more flexible bases, such as **legitimate interests** (Art 6(1)(f)), or on contractual necessity when

[61] GDPR, art 5(1)(a)
[62] GDPR, art 6

the AI processing is closely tied to a service requested by the user. However, these bases come with limitations. Legitimate interest requires a careful **balancing test**: the controller's aims must not be overridden by the individual's rights and interests, "in particular where the data subject is a child".[63] This notice reflects a recognition that certain persons, especially minors or other vulnerable individuals, merit extra protection. Malgieri argues that many data subjects can be vulnerable in the AI context due to power imbalances between individuals and data controllers.[64] For example, an asylum seeker subjected to automated risk profiling or an elderly person using a smart device may not be in a position to meaningfully consent or object. Malgieri's work on vulnerability suggests that GDPR's one-size-fits-all approach to lawful bases may fail to account for these power asymmetries, and that a more nuanced application of lawfulness is needed to protect those who are inherently at a disadvantage.[65] Indeed, Recital 47 GDPR itself notes that children (as a proxy for vulnerable data subjects) will rarely have their interests outweighed by a controller's legitimate interests. [66]

Challenges in Practice: Under the GDPR's Article 6, any processing of personal data must have a specific lawful basis (such as consent, contract, legal obligation, vital interest, public task, or legitimate interests). In theory, this requirement applies fully to AI-powered systems, including those that scrape publicly available data. In practice, however, ensuring a valid legal basis for large-scale AI data processing has proven challenging. AI models often ingest vast amounts of personal information (e.g. images or text from the web) without individuals' knowledge raising the question of how such processing can be lawful, fair, and transparent as required by GDPR principles. The Clearview AI case offers a prime example of these difficulties, as it involves an AI system collecting billions of facial images from the internet and highlights the legal and ethical tensions around Article 6 in an AI context.

[63] Ibid.
[64] Malgieri (n 5) 57-61.
[65] Ibid 167-172.
[66] GDPR, recital 47.

Clearview AI is a facial recognition company that built a database of 3+ billion images by scraping photos of people from publicly accessible websites and social media. It then sold access to this database, mainly to law enforcement, allowing users to identify individuals in a "probe" photo by matching it against Clearview's collection. The personal data involved, including facial photographs, is considered **biometric data,** a special category under GDPR Article 9(1) due to its use in uniquely identifying a person. From a GDPR standpoint, Clearview's activities triggered multiple red flags. European regulators found that Clearview **processed personal data on a massive scale without any valid legal basis under Article 6**. In other words, individuals' photos were collected and used in a new context (facial recognition for law enforcement) without meeting any of the lawful grounds in Article 6(1). Clearview was not established in the EU, but GDPR's **extraterritorial scope** (Article 3(2)) still brought these activities under EU jurisdiction since EU residents' data were targeted. Authorities across Europe, from France's CNIL to Italy's Garante and the Dutch and Greek DPAs, investigated and unanimously deemed Clearview's operations unlawful, ordering the company to stop processing EU data and levying heavy fines. Notably, the Dutch DPA in 2024 imposed a €30.5 million fine, concluding that Clearview's business of scraping and matching faces "**violated the GDPR by processing personal, biometric data without a proper legal basis**". This case vividly illustrates how an AI system built on public data can run afoul of GDPR requirements in practice.[67]

Moreover, even when a **legitimate interest** basis is invoked for AI processing, controllers must document and defend their balancing of interests. In high-stakes AI decisions (e.g. an algorithmic credit score affecting someone's access to loans), the data subject's rights (privacy, non-discrimination, etc.) weigh heavily. Academic commentary suggests that controllers often

[67] Olaf van Haperen, 'GDPR Enforcement Beyond EU-Borders — The Dutch Data Protection Authority's Fine on Clearview AI and the Future of AI Regulation & Enforcement' (2025) 26 *Computer Law Review International* 10, 10-12; Won Kyung Jung and others, 'Privacy and Data Protection Regulations for AI Using Publicly Available Data: Clearview AI Case' in *Proceedings of the 17th International Conference on Theory and Practice of Electronic Governance* (ACM 2024) 48, 48-52.

interpret "necessary for legitimate interests" generously, stretching it to cover large-scale profiling operations. Vrabec observes that while the GDPR formally strengthened individuals' control, in practice the effectiveness of these rights and principles can be limited in complex data-driven systems.[68] Data subjects may not be aware of AI-driven processing to exercise their rights, and legal bases like legitimate interest shift the burden onto individuals to object rather than requiring prior consent. This dynamic can undermine the spirit of the lawfulness principle. Thus, there is a growing call in scholarship for clearer guidance on applying lawful bases to AI, for instance, whether developing certain AI models should be treated as "scientific research" (which enjoys some leeway under GDPR), or how to ensure that vulnerable groups are not exploited under broad legitimate interest claims.[69]

3.1.2 Purpose Limitation (Article 5(1)(b) GDPR)

The principle of purpose limitation requires that personal data be **"collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes."**[70] In other words, when a controller gathers personal data, it must define a clear purpose, and any secondary use of the data should not deviate from the original purpose in a way that data subjects would not reasonably expect. This principle aims to prevent *function creep*, the expansion of data use beyond the context in which the data was first collected. Notably, Article 5(1)(b) includes a built-in accommodation for socially beneficial reuse: further processing for "archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" is **not** considered incompatible with the initial purpose, subject to safeguards per Article 89(1). This exception is highly relevant to AI development, as many AI projects might be characterised as research or statistical in nature. However, outside such contexts, the controller bears the burden to

[68] Bart Custers and others, 'Assessing the Legal and Ethical Impact of Data Reuse: Developing a Tool for Data Reuse Impact Assessments (DRIA)' (2019) 5 *European Data Protection Law Review* 317, 325-330.
[69] Malgieri (n 5) 172-175.
[70] GDPR, art 5(1)(b).

ensure that any reuse of data is either compatible with the original purpose or is carried out under a new legal basis (for example, by obtaining fresh consent for the new purpose or relying on a legal authorisation).

AI Context and Purpose Flexibility: AI systems often thrive on data reuse: repurposing datasets for new algorithms, combining datasets, or using personal data collected in one context to train models for another. This tendency creates tension with the purpose limitation principle. For instance, consider a bank that originally collected customers' data to provide a checking account service. If the bank later decides to aggregate and analyze that data (and maybe external data) to develop a credit scoring AI, this is a new purpose that may not align with the original service purpose. Under GDPR, the bank would need to assess compatibility or seek a new legal basis. Similarly, using photos from social media to train a facial recognition AI intended for law enforcement use is a drastic change of purpose: individuals uploaded photos to share with friends, not to become part of a surveillance tool. Absent the data subjects' consent or a specific legal mandate, such processing will likely breach purpose limitation. Indeed, in the Clearview AI enforcement mentioned above, regulators deemed the reuse of social media images for facial recognition fundamentally incompatible with the original context, reinforcing that broad AI ambitions cannot justify ignoring this principle.[71]

Recognising the friction between rigid purpose limitation and modern AI development, some regulators and scholars have proposed pragmatic approaches. The French CNIL's recent guidance on AI suggests that controllers building **general-purpose AI systems** (which by design may have multiple future uses) can still satisfy purpose specification by describing the **general nature** of the system and its potential applications, even if every downstream use case is not known at the outset. This implies a flexible interpretation of purpose limitation: as long as the AI developer is transparent about the broad category of purposes (e.g. facial recognition for identity verification, or

[71] van Haperen (n 67) 11.

machine learning for credit risk assessment) and implements safeguards, the principle need not paralyse innovation. The CNIL also noted that reusing existing datasets (including those publicly available online) for AI training can be lawful if the new use is **compatible** with the original purpose or if appropriate measures (like anonymisation or additional notice/consent) are in place.

Critiques and Limitations: Despite such guidance, many scholars highlight that purpose limitation in its classical form sits uneasily with the realities of Big Data and AI. Vrabec and others point out that the GDPR's strict rules on secondary use are "*at odds with the reality of the data economy*", where value is often extracted by finding **novel uses** for data beyond those originally envisaged. For example, datasets collected by one company are frequently sold or shared with AI developers to unlock new insights, creating a thriving data trade that GDPR's compatibility test might disallow. Zarsky famously argued that the GDPR is "incompatible in the age of Big Data", specifically calling out purpose limitation as a constraint that discourages beneficial data-driven innovation. The concern is that if every new data use requires either strict continuity of purpose or fresh consent/legal permission, many AI projects could be stifled or pushed underground. On the other hand, privacy advocates defend purpose limitation as a cornerstone of data protection: it forces data controllers to self-discipline their use of personal data and protects individuals from the erosion of context that leads to privacy harms. Vrabec's analysis underscores that without clear purpose boundaries, data subjects lose meaningful control over their data, which might be exploited in ways they never agreed to, undermining trust in digital services.

In response to these challenges, proposals like *Data Reuse Impact Assessments (DRIA)* have emerged. Vrabec (with Custers and Friedewald) has suggested that a **Data Reuse Impact Assessment** framework could help balance the competing interests by systematically evaluating the risks and benefits of repurposing data controllers might identify when a secondary use can be deemed "compatible" or what safeguards (such as additional anonymisation or strict access

controls) could bridge the gap.[72] Such tools aim to inject accountability into data reuse decisions in AI development, ensuring that controllers do not simply disregard purpose limitation but rather justify and mitigate any departure from the original purpose. Ultimately, the principle of purpose limitation stands as a reminder that **context matters** in data processing: AI designers must carefully consider whether the new context of use respects or betrays the original context in which data was collected. If it is the latter, the GDPR would require obtaining new consent or legal authority, a hurdle that, while potentially slowing down development, serves to protect individuals from unseen exploitation of their personal information.

3.1.3 Data Minimisation (Article 5(1)(c) GDPR)

Closely tied to purpose limitation is the principle of **data minimisation**, defined in Article 5(1)(c) GDPR. This principle mandates that personal data must be "**adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed**."[73] In essence, even if data processing is lawful and purpose-bound, controllers should **not collect or retain more personal data than needed** for the specified purpose. Data minimisation operationalises the idea that every additional piece of personal data can represent an added risk to privacy; thus, only the minimum amount of data required to achieve the legitimate purpose should be used. This principle imposes a discipline on AI developers to ask: *Which data are truly necessary to train or operate this model, and which data can we avoid using?* It also links to obligations like storage limitation (not keeping data longer than needed).

[72] Custers, Vrabec and Friedewald (n 68) 324–327.
[73] GDPR, art 5(1)(c).

Minimisation in AI Model Training: In practice, data minimisation can seem at odds with the prevailing ethos of AI research, which often prizes large datasets to improve model performance. Machine learning experts commonly state that "more data" can yield more accurate or robust models. For example, a facial recognition system might perform better with millions of images rather than thousands; a credit scoring algorithm might seek hundreds of data points about each individual to detect subtle predictors of creditworthiness. The challenge is determining what is "necessary" for the purpose. In a credit scoring context, is it necessary to collect social media data or browsing history about a loan applicant, or would a few financial indicators suffice? From the GDPR perspective, the latter is preferable unless the former can be convincingly justified as proportionate and needed. The law expects controllers to calibrate their data collection to the stated purpose, using data protection by design to minimise personal data use. If the same predictive accuracy can be achieved with a smaller feature set or a leaner dataset, then collecting extra personal data would violate the minimisation principle.

Regulators acknowledge that in AI development, a strict interpretation of "minimisation" should not cripple legitimate technical needs. The CNIL's 2025 AI guidance affirmed that the data minimisation principle "**does not prevent the use of large training datasets**" for AI, as long as the data used is **relevant** to the task and **not excessive**. In other words, training an AI on a vast dataset can be consistent with GDPR if the dataset is necessary to achieve acceptable model accuracy or to avoid bias, etc., and if irrelevant data are excluded. The CNIL advises that datasets should be selected and pre-processed (cleaned) to ensure they only contain information useful for the algorithm's stated purpose, thereby avoiding gratuitous personal data that serves no analytical value. For instance, if building a facial recognition AI to distinguish individuals, including images' metadata about location, might be unnecessary (unless the purpose specifically involves location-based identification). Likewise, in a credit scoring model, including sensitive attributes like ethnicity or health information would typically be unnecessary (and likely unlawful under other

provisions) for the purpose of assessing credit risk–removing such data would be part of minimisation and also helps mitigate discrimination.

Critiques and Practical Difficulties: Academic critiques of data minimisation in the AI context highlight a few issues. First, the notion of necessity can be subjective or difficult to evaluate ex ante. AI developers might argue that until an algorithm is tested, it is hard to know which data features are truly necessary; there is a temptation to "collect everything" and let the model find the predictive patterns. This mindset runs counter to GDPR, and scholars like Zarsky have noted that organisations often perceive minimisation as an obstacle to data-driven innovation.[74] If strictly enforced, minimisation could mean foregoing potentially useful data that might improve an AI's functioning or fairness. On the other hand, unchecked data collection raises risks of function creep, bias, and security issues (more data can mean a bigger breach impact). Some research even suggests that overly zealous minimisation might harm fairness: if datasets are pruned without care, they might under-represent certain groups, impairing the model's accuracy for those groups. Thus, there is a nuanced debate about how to apply minimisation without inadvertently affecting the quality and equity of AI outcomes.

Malgieri's perspective on vulnerability adds another layer of critique: when data subjects are in a weak position (due to age, economic disparity, lack of digital literacy, etc.), they might not realise the extent of data being collected about them, or be able to object effectively. This puts the onus on controllers to self-enforce minimisation rigorously for such groups. For example, if an AI recruitment tool asks job applicants (who may be eager for employment and thus vulnerable in context) to submit extensive personal information, those individuals might comply even if much of that information is not strictly required for the hiring decision. It falls to the controller to limit the questionnaire to what is genuinely needed (education, experience, skills) and exclude intrusive

[74] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 The Seton Hall Law Review 2 <https://scholarship.shu.edu/cgi/viewcontent.cgi?article=1606&context=shlr>.

queries (health status, family life, etc.), thereby honouring both minimisation and fairness. Malgieri's work suggests that current data protection law could do more to identify and shield vulnerable data subjects in such scenarios.[75] This could mean interpreting "necessary" in Article 5(1)(c) in light of an individual's vulnerability: if data pertains to a vulnerable person, a stricter necessity threshold might be applied to ensure they are not exploited by excessive data collection. While not explicitly in the GDPR's text, this approach aligns with the Regulation's risk-based ethos (see Recital 75, noting that processing involving vulnerable individuals can indicate higher risk).

In practice, enforcing data minimisation in AI systems requires technical and organisational measures. Techniques like **feature selection, dimensionality reduction, or synthetic data generation** can help limit the use of real personal data. Data protection authorities increasingly expect that controllers will justify each category of personal data used in an AI model (for instance, through the documentation of a Data Protection Impact Assessment). If an AI-driven credit scoring system imports large swathes of personal information, the controller should be prepared to demonstrate why each type of data is **relevant and necessary** to the creditworthiness assessment. In several jurisdictions, we have already seen regulators question companies on why they collected certain data fields. For example, a European bank was reproached for using customers' social media friends list in a creditworthiness algorithm, a practice found to violate minimisation since that data was not essential to the credit decision. Such enforcement examples underscore that the **practical challenge is significant:** controllers must resist the allure of "data maximisation" and instead adopt a careful, evidence-based approach to defining their data needs.

3.1.4 Enforcing Core Principles in AI: Theoretical Intent vs Practical Reality

[75] Malgieri (n 5) 144-149.

The principles of lawful basis, purpose limitation, and data minimisation encapsulate the GDPR's **theoretical intent** to put fundamental rights and individual autonomy at the centre of personal data processing. In theory, an AI developer should, before ever training a model on personal data, lawfully justify the processing, specify a legitimate purpose, and minimise the data used. These requirements are designed to curb the inherent risks of AI-powered decision-making (such as profiling, loss of privacy, and bias) by embedding precautions from the start. When observed, they compel greater transparency and restraint, for example, forcing a company to drop plans to use data in ways that cannot be justified to an average person, or to avoid collecting sensitive data that it has no permission or need to use. The **accountability principle** (Article 5(2) GDPR) further reinforces this by requiring controllers to be able to demonstrate compliance with all the above. In practical reality, however, enforcing these principles in AI systems has proven challenging. AI development is dynamic and complex, often involving multiple data sources, third-party algorithms, and evolving use-cases, a landscape that can outpace traditional compliance checks. Helena Vrabec's research, viewing data subject rights and principles through the lens of the data-driven economy, concludes that there are persistent *gaps between the law's promises and actual practice*.[76] Companies may declare broad purposes in privacy notices (testing the limits of "specified and explicit" purpose) or rely on vague legitimate interests, betting that regulators will struggle to scrutinize technical AI operations. The opacity of AI ("black box" models) can also mask whether minimisation is truly observed e.g., a model might technically use an identifier or attribute in ways not obvious from the input dataset. Furthermore, when AI models generate inferred data or profiles about individuals (such as a prediction of someone's personality or risk level), it raises the question of whether this constitutes a new processing purpose that needs its own legal basis and justification. Many scholars argue that it does, invoking purpose limitation, the profile should only be used for the specific purpose for which it was created, while others point out that GDPR's provisions on profiling

[76] Custers, Vrabec and Friedewald (n 68) 329-330.

(Article 22) and data subject rights (access, rectification, etc.) still struggle to give individuals control over these algorithmic inferences.[77]

Critics like Malgieri advocate for stronger protective measures, especially where AI's impact on vulnerable persons is concerned. He suggests that while the GDPR's principles are sound, their enforcement must be bolstered by a recognition of human vulnerabilities and power imbalances. In practice, this could mean requiring heightened scrutiny or DPIAs for AI systems used on vulnerable populations (e.g. AI in social services or policing), or even disallowing certain data uses outright as a matter of ethics and fundamental rights, something the forthcoming AI Act (AAA) is grappling with on the legislative front. In the meantime, Data Protection Authorities and courts are increasingly active in applying the GDPR principles to AI. Fines and orders against companies like Clearview AI signal that regulators are willing to uphold the lawfulness and purpose limitations strictly. Guidance from bodies like the EDPB and CNIL, as discussed, shows a desire to interpret principles in a way that **accommodates legitimate AI innovation** (through flexibility and context) **without diluting the core protections**.

In conclusion, the GDPR's data-protection principles of lawful basis, purpose limitation, and data minimisation form a crucial framework for governing AI-driven decision-making systems. They encapsulate the theoretical intent to protect personal data from cradle to grave of processing. Yet, as academic critiques by Vrabec and Malgieri highlight, these principles face practical limitations in the era of Big Data and AI. Ensuring meaningful compliance requires continuous effort: controllers must internalise these principles in AI system design (privacy by design), regulators must issue clarifications and enforce the rules in landmark cases, and scholars will continue to debate and refine how these principles can achieve their aims in new technological contexts. The end goal is that even as AI systems become more powerful and pervasive, the fundamental rights of individuals

[77] Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) 162-167 https://doi.org/10.1093/oso/9780198868422.001.0001

to autonomy, privacy, and dignity remain **adequately protected** by the robust application of GDPR's core principles.

3.2 Transparency and Automated Decision-Making: Articles 13–15 and 22 GDPR

Transparency is a core GDPR principle and a precondition for fair, accountable AI. Individuals have the right to be informed about **automated decision-making, including profiling**, that significantly affects them, and to receive meaningful information about the logic involved (GDPR Articles 13(2)(f), 14(2)(g), and 15(1)(h)). In practice, this means that when personal data are used in AI systems to make decisions about people for example, an algorithm determining creditworthiness or job suitability the data subject is entitled to be informed upfront that such automated processing takes place, and upon request, to receive information that helps them understand **how the decision was made** and what it means for them. Articles 13 and 14 GDPR enumerate transparency obligations at the time of data collection, the controller must disclose if it will make **automated decisions producing legal or similarly significant effects**, and provide "meaningful information about the logic involved," as well as the envisaged consequences for the person. Article 15 extends this as an access right: individuals can ask a controller to confirm if they are subject to automated decisions and obtain information about the reasoning. These provisions are often seen as creating a form of "*right to explanation*," though the GDPR does not use that exact phrase. In fact, scholarly debate was sparked by Recital 71 GDPR (which mentions the right to obtain an explanation after such decisions), some academics argued the GDPR implicitly provides an explanation right, while others considered the requirement only to provide generic information about the algorithm's functionality or criteria, not a personalised explanation of the particular decision. Gianclaudio Malgieri and Giovanni Comandé, for example, describe it as a right to "**legibility**" of automated

processing, i.e. the ability to have algorithmic decisions made **comprehensible** in non-technical terms rather than an absolute right to a full rationale in all cases. [78]

Article 22 GDPR is the pivotal provision on automated individual decision-making. Article 22(1) provides that a person "**shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her**." In essence, fully automated decisions that have important impacts (such as denial of a loan, rejection from employment, etc.) are prohibited **unless** an exception applies. The exceptions (Art 22(2)) include when the decision is necessary for entering or performing a contract with the data subject, when it is authorised by law (with suitable safeguards), or when it is based on the data subject's explicit consent. Even when such exceptions permit automated decisions, the controller must implement additional safeguards: at a minimum, the right for the individual to **obtain human intervention**, to express their point of view, and to contest the decision (Art 22(3)). These requirements reflect a legal attempt to preserve human agency and fairness in AI-driven decisions, aligning with the ethical concern that individuals should not be subject to opaque, unchallengeable algorithmic judgments.

Interpreting Article 22's scope has been a matter of legal and scholarly scrutiny. One question has been what constitutes a "decision … based solely on automated processing" and what qualifies as a "similarly significant" effect. If any human is meaningfully involved in the decision, Article 22(1) may not be triggered, which has led to the concern that companies could put a nominal human rubber-stamp on an AI decision to evade the GDPR's strictures. However, recent **case law is clarifying a broad view** of Article 22. In a landmark ruling in 2023 (*C-634/21, SCHUFA Holding*

[78] Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 Intl Data Privacy L 243, arguing that while the GDPR stops short of a full-fledged right to explanation, it guarantees a 'right to obtain an explanation of the decision reached after [automated] assessment' through a combination of transparency provisions and the safeguards of Article 22. See also Bryce Goodman and Seth Flaxman, 'European Union regulations on algorithmic decision-making and a "right to explanation"' (2017) 38 AI Magazine 50, discussing the ambiguity in GDPR's language and the role of Recital 71.

(Scoring)), the Court of Justice of the EU held that the automated calculation of a credit score by a credit agency (which was then used by a bank to decide on a loan) **itself fell under Article 22(1)**. The CJEU decided that when an automated score "draws strongly" on the final decision, the scoring can be considered an automated decision producing significant effects, even if a human (the bank) formally makes the ultimate decision. This expansive interpretation prevents controllers from avoiding accountability by splitting an algorithmic decision-making process between parties. It aligns with the GDPR's protective purpose, as the Court noted, a narrow reading would create a "lacuna in legal protection" if important algorithmic outputs like credit scores were seen as mere preparatory steps.[79] The SCHUFA case also highlighted the tension between **transparency and trade secrets** in AI: the data subject had requested a "detailed account of the logic" of the scoring algorithm, but the credit agency refused, citing its proprietary model as a trade secret. The GDPR itself mandates disclosure of "meaningful information about the logic," but does not force an organisation to divulge source code or algorithms in detail. DPAs and courts have begun to grapple with where the line lies, and individuals need enough information to challenge or understand a decision, but companies argue they must be able to protect intellectual property and prevent gaming of their algorithms. In the *SCHUFA* proceedings, the Advocate General opined (and the Court implicitly agreed) that at least the **essential rationale** of how the score is obtained and used must be provided to the individual. We are seeing a developing jurisprudence that tries to balance these interests: for instance, controllers might fulfil their duty by explaining the categories of data used, the main factors influencing the outcome, and how those factors combine, without revealing the full algorithmic formula. National case law has echoed this, e.g. the French CNIL's enforcement against fintech algorithms and some German decisions insisting on qualitative explanations ("your loan was denied mainly due to your high level of existing debt and past payment incidents") rather than a

[79] *OQ v Land Hessen* (Case C-634/21, CJEU, 7 December 2023) (Schufa Holding AG), interpreting Article 22(1) GDPR broadly to prevent circumvention, holding that an automated credit score transmitted to a third party qualifies as an automated decision if that score is heavily relied upon by the third party (paras 32, 35). The case establishes that the protections of Article 22 cannot be avoided by splitting decision-making between an algorithm and a human intermediary. Notably, the judgment also left open questions about the extent of information that must be provided about the scoring algorithm, an issue tied to Article 15 and Recital 71 GDPR.

mere numeric score or a blank refusal. Academic commentators like Wachter and Mittelstadt have criticised a lack of explicit "right to explanation" in the GDPR, but they concede that, through Articles 13–15 and 22, **a form of explanation right does emerge in practice**.[80]

From a critical perspective, some scholars argue that GDPR transparency may be **necessary but not sufficient** for algorithmic accountability. Even if a data subject is told some information about how an AI system works, they may still find it difficult to meaningfully contest an automated decision, especially given the complexity of modern AI (e.g. deep neural networks defying simple explanation). Malgieri has suggested moving beyond one-off explanations toward continuous accountability: requiring that systems be not just explainable, but **justified** in terms of sound logic and tested for non-discrimination (sometimes phrased as a call for "*algorithmic accountability*" rather than merely transparency).[81] In a similar vein, Floridi notes that between "*soft ethics*" demands for transparency and the "*hard law*" of data protection, GDPR's provisions like Recital 71 imbue an ethical expectation of explanation which is not fully codified as a clear legal right.[82] Nevertheless, enforcement trends show regulators using GDPR to push for more algorithmic transparency. For example, the Dutch courts in the **SyRI case** (2019) invoked human rights and data protection principles to strike down a government profiling system for lacking transparency and risking discrimination, underlining that even if GDPR's text on automated decisions is limited, broader privacy and equality norms can step in.[83]

---

[80] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 76, 78–79; Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI' (2019) *Nature Machine Intelligence* (preprint SSRN, 20 May 2019) https://ssrn.com/abstract=3391293

[81] Gianclaudio Malgieri, '"Just" Algorithms: AI Justification (beyond explanation) in the GDPR' (14 December 2020) *Blog*(on Gianclaudio Malgieri's website) https://www.gianclaudiomalgieri.eu/2020/12/14/just-algorithms/ accessed 26 August 2025.

[82] Luciano Floridi, 'Soft Ethics and the Governance of the Digital' (2018) 31 *Philosophy & Technology* 1, distinguishing between 'soft ethics' (values and principles that inform the spirit of the law) and 'hard law', noting that many ethical principles (eg transparency, fairness) are reflected in GDPR recitals but are not directly enforceable rules. He suggests that GDPR's recitals serve as interpretative guidance imbued with ethical aspirations (what he terms 'soft ethics'), whereas the operative articles may not fully realise those ideals, thereby creating gaps in addressing emerging AI challenges. This divergence underscores the need for additional legal or regulatory measures to bridge the gap between ethical expectations and legal requirements.

[83]Marco Almada, Training Curriculum on AI and Data Protection: Law & Compliance in AI Security & Data Protection (European Data Protection Board, Support Pool of Experts Programme, December 2024) 46–48.

In sum, Articles 13–15 and 22 of the GDPR seek to **lift the veil** on AI-powered decisions and give individuals tools to know when an algorithm is influencing their lives and to object or obtain human review. Case law (like Schufa) and scholarly critiques are actively shaping what "meaningful information" entails, moving the needle toward more substantive explanations rather than mere formal notices. This is an area where legal and ethical principles intersect strongly: the law is enforcing a baseline of transparency and human oversight, while ethical frameworks (discussed in §3.5) often urge even greater openness, intelligibility, and contestability of AI decisions in order to respect individual autonomy and prevent unjust outcomes.

3.3 Risk-Based Regulation in the EU AI Act: Classification and Obligations by Risk Tier

The European Union's **AI Act** (Artificial Intelligence Act) embraces a risk-based regulatory approach for AI systems, in contrast to the GDPR's blanket data protection rules. It sorts AI systems into categories of risk and calibrates obligations accordingly. At the top level are **prohibited AI practices**, uses of AI deemed unacceptable because they pose grave threats to safety or fundamental rights. Article 5 of the AI Act (Council/Parliament compromise text, 2023) provides an explicit list of banned uses. Notable examples include AI systems that deploy subliminal techniques or exploit vulnerabilities of specific groups in ways that **materially distort a person's behaviour** to their detriment (for instance, an AI toy that manipulates a child into dangerous behavior, or algorithms that prey on the elderly's cognitive vulnerabilities to scam them). Another prohibited practice is **social scoring** by public authorities, i.e. ranking individuals' trustworthiness or worth based on AI analysis of their behaviour or personality, as this is viewed as violating human dignity and equal treatment. It is expected to also ban (with narrow exceptions) **real-time remote biometric identification** in public spaces by law enforcement, owing to its intrusiveness and mass

surveillance implications. By categorically outlawing such use cases, the EU affirms that no degree of accuracy or mitigation makes them acceptable: the *lawmaker has decided that certain AI applications are incompatible with EU values*, period.

Below the prohibited tier, the AI Act designates a broad class of "**high-risk AI systems.**" These are defined primarily by their use case. Annexe III of the Act lists areas of application that are considered high-risk when AI is used for those purposes. The list covers AI systems in domains like: recruitment or HR (e.g. CV-scanning algorithms for hiring), education (e.g. scoring of exams), essential private and public services (e.g. creditworthiness assessments that determine access to loans, or algorithms used to allocate social benefits), law enforcement and migration (e.g. predictive policing tools, risk assessment in asylum decisions), and safety components of regulated products (like AI in medical devices or in the operational control of vehicles). If an AI system is intended for any function explicitly listed (and not exempted by a specific clause), it will be classified as high-risk (Art 6). For example, an AI that evaluates loan applications is high-risk because it falls under "credit scoring" in essential services; an AI system for diagnosing diseases is high-risk both as a medical device component and due to its critical impact on health decisions. **High-risk AI systems face the most extensive obligations** under the AI Act. The Act essentially creates a regulatory regime similar to product safety law for these systems. Providers must undergo conformity assessment and compliance procedures before placing a high-risk AI on the market. Key obligations for high-risk AI (set out in Chapter 2 of the Act) include:

Risk Management: Providers must perform an ongoing risk assessment and mitigation process (Article 9) to identify potential harms (to health, safety, fundamental rights) and reduce them to an acceptable level. This echoes the engineering of safety-critical systems, think of it as requiring "AI safety by design."

Data Governance: High-risk AI training data must meet quality criteria (Article 10). The data should be relevant, representative, free of errors where possible, and complete. Critically, Article 10 also demands measures to ensure **bias monitoring and mitigation** in datasets. For instance, if a high-risk AI is used in hiring, the training data should be checked for gender or racial biases, and steps taken to prevent discriminatory outcomes. The AI Act explicitly allows processing of special-category personal data for bias correction in high-risk AI (Article 10 (5) an important interaction with GDPR, as normally such data would be off-limits without consent or another Article 9 GDPR exception). This provision was introduced to ensure that bias in AI can be tackled even if it requires using sensitive attributes (for example, to test if an AI's error rate is higher for a certain ethnic group, one might need to process data about ethnicity in the evaluation phase).

Technical Documentation: Providers must create extensive technical documentation (Article 11) detailing the system's design, purpose, and performance, which can be reviewed by authorities. This documentation should include information on the algorithms used, training methodologies, and testing processes. It serves both accountability and as an audit trail if something goes wrong.

Transparency and User Instructions: Article 13 requires that high-risk AI systems be accompanied by clear instructions for use, and where appropriate, information to users about the AI's capabilities and limitations. If a system might err or has certain conditions under which it should not be relied upon, this must be communicated. High-risk AI that interacts with people (like a recruiting chatbot or a diagnostic assistant) doesn't just have to disclose it's an AI; it also needs to allow human oversight.

Human Oversight: The Act mandates measures for human oversight (Article 14) to ensure that an AI does not operate unchecked in critical decisions. This might involve **human-in-the-loop** mechanisms or training human operators to understand the AI's outputs and intervene when

necessary. For example, a radiology AI tool would be used by medical professionals who can overrule or double-check its assessments, rather than fully automating diagnoses.

Accuracy, Robustness, and Cybersecurity: Article 15 stipulates that high-risk AI systems must meet standards of accuracy and robustness, and be resilient to attempts to tamper with them or to manipulate their outputs. This could translate into requirements for regular model testing, validation on diverse data, and security measures to prevent hacking or adversarial attacks. Providers will need to monitor and patch their AI models as real-world performance issues are discovered (post-market monitoring, Article 61).

These obligations for high-risk AI are enforced through a conformity assessment regime. Some AI systems (especially those as safety components of products) will require **third-party assessment** by notified bodies before deployment; others might allow self-assessment with proper documentation. Once compliant, high-risk AI systems will bear a CE marking to indicate conformity with the AI Act's requirements. Importantly, the AI Act's high-risk framework is a harmonised law: the rules will be directly applicable across all EU member states, preventing divergent national standards.[84] Providers and deployers of high-risk AI are subject to oversight by national Market Surveillance Authorities, analogous to how product regulators oversee the safety of machinery or toys.[85] The Act thus brings an ex ante compliance approach to AI, complementing the ex post, rights-driven approach of the GDPR.

Next, the Act identifies some "limited risk" AI systems, which are generally allowed but subject to specific transparency obligations. These include AI systems that interact with humans, emotion recognition systems, and AI that generates or manipulates content (so-called "deepfakes"). For example, a chatbot or voice assistant that a person might think is human must disclose that it is AI

[84] Ibid 68-70.
[85] Ibid 71.

Article 52 (Transparency obligations) requires that users are informed they are engaging with a machine, so as to avoid deception 61 62. Similarly, an AI system that creates deepfake images or videos (e.g. an app that fabricates realistic faces or voices) must clearly label the output as AI-generated unless used for legitimate law enforcement or journalistic purposes. These obligations aim to mitigate manipulation and inform individuals, but they are relatively light-touch; there is no need for conformity assessments or audits for limited-risk AI. The Act essentially says: as long as these AI are not in the high-risk categories, we trust general laws and minimal transparency to handle them. Many common applications (AI in video games, spam filters, price optimisation tools, etc.) will likely fall into minimal or limited risk classes and thus face no AI Act requirements beyond existing consumer protection or sectoral laws. The Act explicitly states that AI systems other than those covered as prohibited or high-risk are largely left to existing regulations and voluntary codes.

A notable innovation late in the legislative process was the introduction of rules for **General Purpose AI (GPAI)**, such as large language models (e.g. GPT-style models) that can be adapted to a wide range of uses. Because the risk-based approach "by purpose" doesn't neatly fit these multi-use technologies, the final EU AI Act imposes some baseline obligations on GPAI providers across the board. Providers of general-purpose AI models must, for instance, **comply with EU data governance standards (including copyright)** and **publish basic information** about the model (Article 52a in Parliament text, roughly Article 53 in the consolidated version). They have to disclose things like the model's intended functions, the training data sources (at a high level), and known limitations. Furthermore, if a GPAI model is of a certain scale (meeting defined thresholds of computational resources used, etc.), it may be designated as having **"systemic risk"** and incur additional requirements. For example, a very advanced general model that could be repurposed for critical tasks might be required to implement even more stringent risk mitigation and to **register in an EU database**. These GPAI rules reflect concern about foundational models that could have

far-ranging impacts, rather than trying to classify them as high-risk for each possible use, the Act establishes horizontal duties for their developers (like OpenAI or Google). This is a moving area, but it shows the EU's intent to future-proof the legislation to some extent, so that new AI paradigms (e.g. GPT-4, GPT-5) are brought into the regulatory fold.

To illustrate the **differences in obligations by risk tier:** consider facial recognition technology. If used by police in real-time on street cameras (live biometric ID of pedestrians), that's generally **prohibited** (unacceptable risk, barring narrow exceptions like searching for a specific missing child or imminent terror threat). If the same technology is used in a more contained high-risk scenario say, an employer uses facial recognition for secure access to offices it might be classified as high-risk (biometric identification for access control is in Annex III), requiring the company and provider to comply with the whole suite of requirements (bias testing to ensure it doesn't discriminate by ethnicity, transparency to employees, human oversight in case of failures, etc.). If facial recognition is used just to unlock your smartphone (for personal use, not a mass deployment affecting others), it likely falls under minimal risk, the Act would not impose specific requirements, relying on product safety standards and GDPR for any personal data issues. Finally, an AI-driven face filter on a social media app (that alters your appearance in photos) could be a limited-risk case: it should disclose that the image is altered (to avoid deepfake misuse), but otherwise it's not heavily regulated by the Act.

The risk-based model of the AI Act has been praised as a **proportionate approach** focusing regulatory burden where the stakes are highest but also faces critique. Some scholars argue the lines are blurry: a seemingly innocuous AI system can become high-risk if repurposed, and some high-risk uses might slip under thresholds. There is also the matter of enforcement: the Act will rely on national agencies (some possibly data protection authorities, or new AI authorities) to monitor compliance, which may be challenging given technical complexity. Nonetheless, the Act marks a

significant expansion of EU tech regulation beyond data protection, aiming for comprehensive AI governance. It consciously complements the GDPR: as the EDPB has noted, the AI Act and GDPR will work in tandem, the AI Act addressing **ethical and safety risks of AI as such**, and the GDPR continuing to safeguard personal data and privacy.[86] For instance, both laws deal with bias: the AI Act demands bias mitigation in model design, while the GDPR can address bias through its fairness and non-discrimination principles (and via rights like rectification or objection if biased personal data is used). The AI Act's concept of "trustworthy AI" operationalises many ethical principles (transparency, accountability, etc.) into concrete requirements for high-risk systems, which we discuss further in Section 3.5. In summary, the EU AI Act stratifies AI regulation into tiers of risk, banning the worst, strictly controlling the high stakes, informing about the lesser risks, and largely tolerating the benign. With this hierarchy, the EU seeks to prevent harmful AI uses while not stifling benign innovation. How well these lines hold in practice will depend on clear guidance and vigilant enforcement, but the approach itself underscores the principle that the level of governance should match the level of potential harm or impact of an AI system.

3.4 Impact Assessments Under the U.S. Algorithmic Accountability Act (AAA)

Across the Atlantic, lawmakers have proposed the **Algorithmic Accountability Act (AAA)** to address AI-driven decision systems. The AAA (a bill introduced in the US Congress, reintroduced in 2025) takes a different regulatory tack, rather than prescribing substantive rules for all AI; it focuses on procedural accountability via **algorithmic impact assessments**. The Act would empower the Federal Trade Commission (FTC) to require that covered entities (generally, large companies or those handling large amounts of personal data) conduct algorithmic impact assessments for their automated decision systems and "augmented critical decision processes."[87] In

[86] Ibid 72-76.

[87] Algorithmic Accountability Act of 2025, S.2164, 119th Congress (2025) §2(1)–(7) – defines "automated decision system" broadly as any software or process that uses computational techniques (including AI or machine learning) to make a decision or aid decision-making. "Augmented critical decision process" refers to a decision process in a critical domain (healthcare, finance, employment, etc.) where an automated system is used to inform or make the decision. The

plain terms, if a company uses AI to make important decisions about consumers, such as algorithms deciding on loans, insurance, hiring, or content moderation, it must regularly audit and evaluate those systems for a variety of risks and document how those risks are being addressed.

Under the AAA, a "**covered entity**" (for example, a tech company above a certain revenue or data size threshold) deploying an automated decision system in contexts that significantly impact consumers (termed "critical decisions") would be obligated to perform an impact assessment of that system. The Act outlines in detail what an **impact assessment** must entail, reflecting a comprehensive checklist of ethical and governance considerations. Key components of the required **Algorithmic Impact Assessments (AIA)** include:

- **Privacy and Data Protection Evaluation:** Companies must examine how the automated system handles personal data and what privacy risks it poses. This involves assessing data minimisation practices and retention limits e.g. documenting what personal data the AI ingests, how long it keeps data or output decisions, and whether data not strictly necessary for the algorithm's purpose is being purged. The AAA expects entities to implement **privacy-enhancing measures** (like de-identification, encryption, or federated learning) and to account for data security. An assessment should thus enumerate steps like differential privacy or other safeguards used to protect individuals' information. If an AI is making sensitive decisions, the company should be able to show it has evaluated risks of data leaks or misuse and taken steps to mitigate them (for instance, preventing a hiring algorithm from inadvertently exposing applicants' health data).

- **Algorithmic Fairness and Non-Discrimination Testing:** The AAA places significant emphasis on **testing for bias** and disparate impacts. Impact assessments must include **"ongoing testing and evaluation of the performance"** of the AI system, with attention to how outcomes may differ

Act's scope of "covered entities" targets larger companies (exceeding revenue or data processing thresholds) deploying these systems.

across **protected or sensitive classes** (race, gender, age, etc.). Concretely, a company might be required to use benchmark datasets or its own historical data to see if, say, a lending algorithm disproportionately rejects borrowers of a certain ethnicity or if a hiring algorithm systematically scores female candidates lower. The Act lists protected characteristics (race, colour, sex, gender, age, disability, religion, socioeconomic status, etc.) and insists on documenting the **methodology** for evaluating differential performance, including what proxies (like ZIP codes for race) were used to detect bias. If subpopulations are tested separately, the assessment should explain which groups were chosen and why. This essentially mandates a **fairness audit o**f AI, a practice already emerging in industry, but the AAA would make it a legal duty. The findings (e.g. that an AI has higher error rates for certain groups) would then presumably force the company to adjust the system or justify its design. The AAA even requires that the assessment document any **limitations on use** that were considered to prevent harm, for instance, if an AI is deemed too biased for a certain application, the company might restrict that use or build in manual checks.

- **Stakeholder and Impacted Community Consultation:** Uniquely, the AAA directs companies to **consult stakeholders** both internal (like ethics teams, affected employees) and independent external stakeholders (such as civil rights advocates or domain experts) as part of the impact assessment process. The Act requires documentation of these consultations: who was consulted, when, and what feedback they provided. It even asks for notes on whether stakeholders' recommendations were incorporated or not, and if not, an explanation of why. This is a significant attempt to inject *democratic accountability* into AI governance: affected communities or experts should have a voice in evaluating an algorithm's impact. For example, a company deploying a facial recognition system in job interviews might consult with privacy groups or diversity advocates; their input (say, concerns about accuracy for darker-skinned candidates) would become part of the official assessment record. This provision aligns with the ethical principle of **inclusive**

governance, recognising that those who bear the risks of AI should have some say in its design or deployment.

- **Transparency and Documentation:** The AAA heavily emphasises **record-keeping** and transparency of the assessment itself. Companies must *document their findings and actions* in a detailed report. This includes documenting the AI system's design purpose, how it was evaluated, what data it uses, and any measures to mitigate risks. The Act specifies that the documentation should cover the data sources and **origins of the data**, including whether data was collected directly, inferred, purchased, etc., and whether individuals gave consent for their data to be used in that way. For instance, if a system uses consumer data scraped from social media, the report should state that and note if users had consented to such use. Additionally, any limitations of the algorithm identified during assessment or any "**ethical guardrails**" considered (e.g. deciding not to use the AI for certain high-stakes decisions) should be recorded. The AAA envisions that these impact assessment reports (or at least summary findings) could be made available to regulators and perhaps the public. Indeed, prior versions of the bill suggested that companies would have to publish an **executive summary** of their AIA to foster public accountability. The 2025 bill text indicates the FTC will set formats for summary reports, implying that some information will become public-facing (except for trade secrets or confidential business info, which the Act protects from disclosure).

- **Ongoing Monitoring and Education**: Recognising that algorithms and their contexts change, the AAA frames impact assessment not as a one-off checkbox but as an **ongoing process**. It requires periodic reassessment, especially when an AI system is updated or when its use changes significantly. The Act also interestingly calls for companies to **train and educate their staff** about the AI's known issues and best practices. For example, if an impact assessment revealed that a hiring algorithm has certain blind spots, the HR team and engineers should be informed and

trained on how to interpret or improve the model. Furthermore, companies must stay abreast of emerging best practices the Act encourages consultation of "relevant proposals and publications" by experts, meaning firms should continuously incorporate the state-of-the-art in ethical AI (e.g. new bias mitigation techniques).

The scope of the AAA is notable it doesn't try to ban specific AI uses (as the EU does) but rather to force *self-scrutiny* and transparency. The logic is that through mandated audits, problems like bias or privacy risks will be surfaced and can then be addressed either by the company or through regulatory action if not fixed. It is analogous to how environmental law requires an environmental impact assessment for big projects: not outright forbidding them, but compelling due diligence and public disclosure of risks. If the AAA is enacted and enforced by the FTC, companies would likely have to file these assessments and could face penalties for non-compliance or for issues like bias that are not mitigated.

Critics of the AAA approach point out that its effectiveness will depend on FTC enforcement muscle and the thoroughness of the assessments. If companies conduct only perfunctory analyses or bury unfavorable findings, the goal of algorithmic accountability might not be achieved. However, the very process of requiring multidisciplinary input and documentation can create internal pressure for better practices (e.g. engineers knowing their work will be examined for fairness may take more care from the start). The AAA also complements anti-discrimination law: whereas existing US laws (like the Equal Credit Opportunity Act or employment discrimination law) prohibit disparate impact based on protected traits, they often don't mandate proactive bias testing. The AAA would fill that gap by ensuring companies look for disparate impacts before they cause widespread harm.

In summary, the Algorithmic Accountability Act's regime of **algorithmic impact assessments** seeks to operationalise ethical AI principles of fairness, transparency, accountability, and respect for

privacy through a legally mandated audit process. It reflects an understanding that AI ethics cannot be left solely to corporate voluntarism; rather, companies should "show their work" when they deploy algorithms that shape important decisions about individuals' lives. This focus on process and documentation diverges from the EU's prescriptive rules, but it is another path to the same end: ensuring *AI-powered decision-making is subject to oversight, explanation, and improvement.* The AAA has not (as of 2025) been passed into law, but its introduction (with bipartisan support in some cases) has already incentivised some companies to begin adopting internal algorithmic audit practices in anticipation of regulatory requirements. Should it pass, we would expect a new compliance landscape in the US where **algorithmic accountability reports** become as routine (and as vital) as financial audits for major corporations deploying AI.

3.5 Ethical Frameworks: Fairness, Non-Discrimination, Accountability, Autonomy, and Human Dignity

Legal mandates like the GDPR, the EU AI Act, and the AAA are increasingly influenced by and interwoven with **ethical frameworks for AI** that have emerged internationally. These frameworks articulate high-level *normative principles* to guide the development and deployment of AI in a way that upholds human values. Several notable examples include the **UNESCO Recommendation on the Ethics of AI (2021)**, the **OECD AI Principles (2019)**, and the EU's own **Ethics Guidelines for Trustworthy AI (2019),** developed by the High-Level Expert Group. Despite originating from different bodies, these instruments converge on a set of core principles among which *fairness*, *non-discrimination*, *transparency/explainability*, *accountability*, *human autonomy*, and *human dignity* feature prominently.[88]

[88] **OECD**, *Recommendation of the Council on Artificial Intelligence* (adopted May 2024; revising the 2019 OECD AI Principles), the OECD AI Principles—five values-based principles and five policy recommendations on innovative, trustworthy AI that respects human rights and democratic values.

For instance, UNESCO's Recommendation (a global agreement endorsed by 193 countries) is grounded in the "protection of **human rights and human dignity**," and it emphasises principles such as **fairness**, **transparency**, and **human oversight** of AI systems.[89] It calls on AI actors to promote social justice and ensure that AI **does not perpetuate bias or discrimination**, reflecting a commitment to equality and fairness internationally. The UNESCO Recommendation also stresses that humans must retain control: AI should be subject to **human oversight** and never be allowed to undermine human autonomy or agency. Similarly, the **OECD AI Principles** which have been adopted by dozens of countries (including the US and EU members) set five value-based tenets: (1) AI should benefit people and the planet by driving **inclusive growth, sustainable development and well-being**; (2) AI systems should respect the rule of law, human rights, and **democratic values**, which include principles of **diversity, non-discrimination, fairness**, and *justice*; (3) they should ensure **transparency** and **explainability** so that people can understand AI outcomes; (4) they must be **robust, secure, and safe** throughout their life cycle; and (5) there should be **accountability** for AI systems, meaning someone is responsible for the AI's behavior and outcomes.[90] These principles, though non-binding, have informed national AI strategies and indeed the drafting of the EU AI Act (one can see direct echoes, e.g., the AI Act's focus on safety and accountability, and its bias mitigation requirements, mirror the fairness and safety elements of OECD's framework).

The EU's **Ethics Guidelines for Trustworthy AI** (2019) distilled similar ideas into a comprehensive concept of "Trustworthy AI" built on three pillars: it should be lawful, ethical, and robust. The guidelines enumerate **seven key requirements** for AI: (1) **Human agency and oversight** AI should empower people and not diminish their autonomy; there should always be the possibility for human intervention or control. (2) **Technical robustness and safety** AI should be

[89] *Recommendation on the Ethics of Artificial Intelligence* (UNESCO General Conference, adopted November 2021), the first global normative instrument on AI ethics encompassing values such as transparency, fairness, human oversight, and interconnectedness across 11 policy action areas.
[90] OECD, Recommendation on Artificial Intelligence (n 86)

secure and reliable. (3) **Privacy and data governance** AI must respect privacy and ensure quality data management. (4) **Transparency** encompassing explainability, traceability, and open communication about AI systems' capabilities and limitations. (5) **Diversity, non-discrimination and fairness** AI should avoid bias and be accessible, and it should foster equal opportunity. (6) **Societal and environmental well-being** AI should benefit society at large, supporting sustainability and social responsibility. (7) **Accountability** mechanisms should be in place to assign responsibility and ensure compliance with the above (for example, auditability, the ability to report problems, redress for adverse impacts). These guidelines were advisory for industry and policy, but they had a direct influence on later policy: for example, the AI Act's provisions on human oversight and transparency trace back to the HLEG's first principle. They also explicitly tie into fundamental rights, fairness is linked to non-discrimination rights, explicability and accountability link to justice, bridging ethical values with legal norms.[91]

The **principle of fairness and non-discrimination** is a linchpin across ethical codes. Ethically, fairness entails both *procedural fairness* (transparent and impartial processes) and *substantive fairness* (outcomes that are equitable and do not create unjustifiable disparities). International instruments (like the OECD and UNESCO texts) urge that AI be designed to **avoid perpetuating biases** and to promote *inclusive* outcomes. They emphasise vigilance against AI that could **exacerbate existing inequalities or create new ones** for instance, AI used in criminal justice that disproportionately affects minority groups would violate these principles. While non-discrimination is also a legal principle (indeed, equality is enshrined in many constitutions and laws), ethical frameworks often go further: they call for proactive fairness audits and inclusion of diverse perspectives in design, not just reactive legal compliance. For example, UNESCO's principle of fairness and non-discrimination implies that AI practitioners should actively work to eliminate dataset bias and ensure algorithms treat different populations justly. The EU's ethics guidelines even

[91] *Ethics Guidelines for Trustworthy AI* (High-Level Expert Group on Artificial Intelligence, European Commission, April 2019), published 8 April 2019, outlining seven key requirements for trustworthy AI such as human oversight, robustness, transparency, fairness, and accountability.

mention a *"commitment to equal and just distribution of benefits and risks"*, tying AI fairness to ideas of social justice (solidarity) and requiring that AI's advantages do not only accrue to the powerful. This philosophical notion of fairness can be broader than what any single law covers for instance, *"fairness"* might include fairness between generations (AI's impact on future jobs) or fairness to non-human life (environmental impact), aspects not captured in current legal statutes.

Accountability is another cross-cutting ethical principle. It demands that organisations and individuals developing or deploying AI be responsible for its outcomes; they should be able to explain and justify their AI's behaviour, and mechanisms should exist to remedy harm. The OECD principle explicitly lists accountability, and the EU ethics guidelines devote a requirement to it, linking it with fairness (an accountable AI actor will take responsibility if something unfair occurs). Accountability in ethical terms can mean setting up governance structures like ethics boards, algorithmic audit trails, and routes for external scrutiny. Legally, GDPR's **"accountability"** (Art 5(2)) requires data controllers to demonstrate compliance (more inward-facing), whereas AI ethics calls for a broader accountability: including to the public, to stakeholders, and even considering moral accountability (who *should* be answerable if an autonomous system causes harm?). Luciano Floridi and Josh Cowls, in analyzing myriad AI ethics charters, identified "**explicability**" as a new, necessary principle comprising both **intelligibility** (transparency, explainability) and **accountability** (answerability for decisions).[92] Their point was that ethical AI requires not just that we can understand how it works, but also that there is someone to answer the question **"who is responsible for the way it works?"**. In practice, this has spurred ideas like *AI ethics boards*, *algorithmic accountability reports* (as we see with the AAA), and even discussions of *legal personality for AI* (though that idea has been largely set aside, with focus returning to holding the humans behind AI accountable).

[92] Ibid.

Autonomy and human dignity are perhaps the most profound ethical principles. *Human autonomy* means that AI should not undermine people's ability to make free choices and act on their own values. Ethically, this underpins warnings against AI practices like manipulative targeted advertising, *"dark patterns"* in user interfaces, or social media algorithms that exploit psychological biases to drive engagement at the cost of user well-being. The EU's ethics guidelines make "respect for human autonomy" the first principle, emphasising that AI should be designed to **augment human abilities, not coerce or deceive**. They caution that humans must retain self-determination AI should not become a form of automated *"nudge"* that irresistibly shapes one's decisions without consent. This ties closely to **human dignity**: the inherent worth of each person and the right to be treated as an end in themselves, not merely as a data point or means for an AI's function. UNESCO's declaration that human rights and dignity are the cornerstone implies that, for example, AI systems should not humiliate, degrade, or reduce people to stereotypes. A dignified treatment also means allowing for human contact and review in sensitive matters one might say Article 22 GDPR, by requiring a possibility of human involvement in high-impact automated decisions, resonates with a dignity concern (people should not be subjected to a computer's verdict with no human consideration). There is also an angle of **autonomy in the sense of consent and control**: ethically, individuals should have agency over whether and how AI systems affect them. This is reflected in calls for robust consent mechanisms, the ability to opt out of AI-driven processes, or at least to contest and correct them (concepts like a right to an explanation or right to a human review have roots in both ethics and data protection law).

Another principle often mentioned is *beneficence* or **"do good"** AI should ideally promote well-being and social good. Luciano Floridi et al. included *beneficence* (promoting well-being) and *non-maleficence* (avoiding harm) as two of the five core principles in their framework.[93] This is mirrored by UNESCO's emphasis on environment and society AI ethics isn't only about avoiding

[93] Luciano Floridi and Josh Cowls, 'A Unified Framework of Five Principles for AI in Society' in Luciano Floridi (ed), *Ethics, Governance, and Policies in Artificial Intelligence* (Springer 2021) 5.

harm, but also about actively using AI to *positively impact society and the environment*. For example, AI should help achieve sustainable development goals, not just make profitable decisions. While our thesis chapter focuses on governance of decision-making, it's worth noting that ethical frameworks broaden the horizon to ask, *"Is this AI use truly good for humanity?",* a question that goes beyond compliance. In some cases, this leads to recommendations to **avoid certain AI altogether** (echoing the prohibitions we see in the AI Act, such as social scoring, which offends human dignity). In others, it means encouraging AI for beneficial uses (like AI for healthcare or accessibility) and ensuring equitable access to AI's benefits across different communities (tying back to fairness).

It's important to highlight **divergences between these ethical ideals and legal implementations**. Often, ethics guidelines are more aspirational and wide-ranging than the law. For instance, ethics may demand **bias-free AI** in a broad sense of justice, but the law (like GDPR or nondiscrimination statutes) might only prohibit *specific kinds* of bias (e.g. discrimination against protected classes in specific contexts such as employment or credit). An AI system could therefore be *legally compliant* (not overtly violating any discrimination law or privacy law) yet still be judged *unethical* by broader criteria, perhaps it unfairly impacts the poor (socioeconomic status not being a protected category in some jurisdictions) or undermines user autonomy in subtle ways. We saw an example in the discussion of consent: GDPR might consider a consent valid with a checkbox click, but an ethical view (esp. from a behavioural perspective like Vrabec's research) might argue that users are often manipulated or not truly free, so ethically the standard for respecting autonomy should be higher than the bare legal requirement. Another divergence is **transparency**: laws like GDPR require information disclosure, but they don't necessarily ensure *comprehension*. The ethical principle of transparency would imply that AI decisions and systems should not only be disclosed but also understandable to those affected (hence the push for explainable AI). In practice, many GDPR privacy notices about AI are written in complex legalese, formally satisfying the

transparency duty but arguably failing the ethical spirit of enabling an average person to genuinely understand the algorithm's logic.

However, ethical frameworks and legal norms are increasingly informing each other. The EU's legislative initiatives explicitly reference ethical principles (the AI Act's preamble cites the goal of trustworthy AI and the work of the AI HLEG). Likewise, soft law like ethics guidelines often recommend **legal or regulatory changes** when voluntary adoption seems insufficient for example, the Ethics Guidelines for Trustworthy AI led to a piloting process and implicitly paved the way for the binding AI Act. Scholars such as Floridi have noted that **ethics can precede law as a "sense-making" exercise**, identifying new issues and values at stake, which law can then codify ("hard ethics"). The flip side is also true: law can reinforce ethics by providing concrete enforcement. For instance, bias mitigation moved from an ethical ask to a legal must (in high-risk AI systems under Article 10(3) of the AI Act). [94]

To draw on the requested perspectives: Helena Vrabec's work has often focused on the notion of *individual control* and the psychological reality behind data subject rights.[95] From an ethical standpoint, her perspective suggests that empowering individuals (autonomy) requires understanding human behaviour; simply giving rights (as GDPR does) may not suffice if users face "consent fatigue" or dark patterns. Ethically, that implies regulators should perhaps require more user-friendly, engaging transparency and genuine choice architectures, not just tick-box compliance. Malgieri's scholarship (apart from legal analyses) frequently emphasises substantive fairness and vulnerability ethically; he posits that AI governance should pay special attention to vulnerable groups and power imbalances.[96] This has echoes in the ethical principle of justice (to each their due, and protecting the weak). Legally, this might translate to stronger protections or enforcement in

[94] Almada, *Training Curriculum on AI and Data Protection* (n 83) 63-65.
[95] Vrabec HU, *Data Subject Rights under the GDPR* (OUP 2021) 35-41 https://doi.org/10.1093/oso/9780198868422.001.0001
[96] Malgieri (n 5) 61-66.

contexts affecting vulnerable populations (as he advocates, e.g., a DPIA focusing on vulnerable data subjects, or DPAs using their powers to guard against exploitative AI). Luciano Floridi's philosophical contributions underscore the need for an **ethical framework that is both actionable and grounded in human values**, like the five principles he and Cowls outline. Floridi often remarks that ethical AI isn't just about avoiding what's illegal, but about doing what is *"good",* something that law alone cannot guarantee. He also warns against "ethics washing" (invoking ethical principles in words but not in deeds or enforcement). Hence, a divergence sometimes exists between high-level ethical proclamations and actual legal requirements or corporate behaviour.

In conclusion, integrating ethical principles with legal governance is crucial for AI-powered decision-making systems. Where law (like GDPR or the AI Act) provides the **minimum baseline** prohibiting clear harms, requiring transparency, and guarding rights, ethics provides the **aspirational benchmark** for truly trustworthy AI, promoting human flourishing, dignity, and social justice. Divergences arise when compliance with the letter of the law does not fully achieve the spirit of ethical AI. Bridging that gap is an active endeavour: regulators issue guidelines, standards bodies develop technical norms for fairness or explainability, and companies adopt internal ethics charters. As AI continues to evolve, the dialogue between ethical frameworks and legal rules will likely intensify. We may see more ethical principles being "hardened" into law (as with bias audits or transparency requirements), and conversely, legal boundaries informing what is considered ethical (for example, if certain AI uses are outlawed, they become ipso facto unethical for companies to pursue). Ultimately, effective AI governance will depend on both a strong legal/regulatory backbone *and* a robust ethical culture, ensuring that AI systems not only comply with rules but actively **embody the values** that our societies cherish, fairness, autonomy, dignity, and accountability chief among them.

4. Comparative Analysis of GDPR, the US Algorithmic Accountability Act, and the EU AI Act

The GDPR, the EU AI Act, and the US Algorithmic Accountability Act (AAA) differ markedly in scope and approach. This section compares their geographic reach, regulatory style, remedial mechanisms, and the cross-border compliance challenges they pose. The analysis draws on primary texts and commentary to highlight where the frameworks converge (e.g. impact assessment requirements) and diverge (e.g. binding vs discretionary duties).

4.1 Regulatory Scope

Extraterritorial Reach

Both the GDPR and the EU's AI Act employ an **extraterritorial scope**, extending their reach beyond Europe's borders. **GDPR** applies not only to EU-based entities but also to any organisation worldwide that processes personal data in the context of offering goods/services to individuals in the EU or monitoring their behaviour.[97] This broad territorial scope means even non-European companies can fall under GDPR obligations if they handle EU personal data. Similarly, the **AI Act** will capture providers and users of AI systems outside the EU whenever their AI systems' outputs are used in the EU or they place AI products into the EU market. In other words, a North American or Asian tech company deploying an AI service that is accessible to Europeans must comply with the AI Act's requirements – a deliberate design to prevent jurisdictional loopholes. This mirrors the "Brussels Effect," whereby EU tech regulations set de facto global standards due to the EU market's size and strict rules.[98]

[97] OJ Gstrein and AJ Zwitter, 'Extraterritorial Application of the GDPR: Promoting European Values or Power?' (2021) 10 *Internet Policy Review s 2.1* https://doi.org/10.14763/2021.3.1576
[98] Ibid s 2.1 and s 3.

By contrast, the **AAA** has a narrower jurisdictional ambit. Earlier iterations of the U.S. Algorithmic Accountability Act (2019; 2022) were relatively narrow in scope, focusing on mandatory impact assessments for automated decision systems (ADS) deployed by large companies, with obligations limited to systems affecting privacy, fairness, or security in critical decision-making contexts. Enforcement was delegated to the Federal Trade Commission, but sectoral exemptions (e.g. banking, non-profits, common carriers) remained intact, and the assessment requirements lacked specificity.

The 2025 proposal builds on these foundations but significantly extends them. In addition to the size thresholds of earlier drafts, it captures smaller entities (>$5M annual revenue) whose systems are integrated into larger covered entities' operations. It also introduces the novel concept of **Augmented Critical Decision Processes (ACDPs)**, explicitly targeting algorithmic tools that shape consequential human decisions. New obligations include detailed **Algorithmic Impact Assessments** encompassing governance, stakeholder consultation, bias detection, explainability, and public disclosure via an FTC repository. Thus, whereas earlier versions emphasised internal oversight, the 2025 Act foregrounds t**ransparency, external accountability, and participatory governance,** marking a substantive evolution in U.S. legislative approaches to algorithmic regulation.

Overall, the GDPR and the EU AI Act assert a form of **global regulatory reach**, applying to any AI-driven personal data processing or AI system that concerns individuals in the EU, **regardless of where the controller or provider is established.** By contrast, the U.S. Algorithmic Accountability Act remains a **domestically oriented framework**, confined to entities under the jurisdiction of the **Federal Trade Commission**. Its obligations chiefly bind large private-sector actors operating within U.S. commerce, without the express extraterritorial effect characteristic of EU instruments.

4.2 Regulatory Approach

The EU and U.S. frameworks represent fundamentally different regulatory philosophies for AI governance.

The EU frameworks are comprehensive and prescriptive, whereas the AAA is risk-management-oriented and procedural. Both the GDPR and the AI Act are **binding EU Regulations** enacted by the legislature. They set out detailed legal requirements and enforceable duties with stiff penalties for non-compliance that apply uniformly across the Member States. Under the **GDPR**, regulation of automated decision-making is embedded in a broad data protection regime: controllers must follow principles like fairness, transparency, data minimisation, and accountability for all personal data processing. Specific to AI, GDPR Article 22 restricts fully automated individual decisions that have legal or similarly significant effects, unless certain conditions are met, and it mandates safeguards such as the right to human intervention.

The **AI Act** builds on this rights-centric approach with a **four-tier risk classification** for AI systems from minimal risk to unacceptable risk. **High-risk AI systems** must undergo mandatory conformity assessments, rigorous data governance, human-oversight measures, and post-market monitoring. All of these obligations from technical documentation (Art. 11) to mandatory risk management systems are spelt out in the text of the Act and will be enforced by EU authorities, and quality controls before placing such systems on the market. It outright **prohibits "unacceptable risk" AI** uses (e.g. social scoring systems or manipulative AI that exploits vulnerabilities). Similarly, the GDPR lays down numerous substantive requirements, such as data-minimisation and DPIAs under Articles 5 and 35, with penalties up to €20 million or 4% of global turnover for breach.

By contrast, the **AAA's model is less prescriptive on substantive design rules**, focusing instead on procedural accountability through **algorithmic impact assessments**. The AAA (as introduced in 2025) would **require covered entities to evaluate their automated decision systems for potential harms** (bias, discrimination, privacy and security risks, etc.) **both before deployment and periodically after**.[99] In essence, it imports the concept of **ex ante and ex post impact assessments**, a governance tool long used in Europe via Data Protection Impact Assessments and now mandated by the AI Act's conformity assessment and post-market monitoring into U.S. law.[100] Each covered AI system or "augmented critical decision process" must undergo a documented review of its training data, performance, and impacts on accuracy and fairness. **Rather than forbidding specific AI use-cases**, the AAA's aim is to ensure organisations **identify, mitigate, and transparently document risks** in their AI systems. This reflects a **"process-and-documentation" regulatory style**: the law mandates *procedures* (risk assessments, transparency reports) and internal governance structures but leaves many implementation details to the regulated entities and further rulemaking. Indeed, the AAA explicitly delegates significant authority to the Federal Trade Commission (FTC) to **promulgate rules and technical standards** for these algorithmic assessments.[101] This reliance on **agency rulemaking** means the *concrete* obligations under AAA would be fleshed out over time by the FTC in contrast to the AI Act's very granular obligations set by the legislature. In short, the EU's model establishes **clear upfront requirements and categorical prohibitions** (a form of "hard law" regulation for all high-risk AI), whereas the AAA embodies a **flexible, adaptive framework** that asks companies to "show their work" in managing AI risks, under oversight of a regulator. The AAA's approach aligns with broader U.S. regulatory trends favoring **ex post accountability and sector-specific guidance** over sweeping ex ante rules. One advantage of this model is agility, the FTC can update guidance as technology evolves but a noted downside is **lack of specificity** and legal certainty: the AAA is relatively high-level and leaves crucial aspects (e.g. what an acceptable

[99] Jakob Mökander, Prathm Juneja, David S Watson and Luciano Floridi, 'The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?' (2022) 32 *Minds and Machines* 751, 758–760 https://doi.org/10.1007/s11023-022-09612-y
[100] Ibid 760-762.
[101] Ibid 762-764.

impact assessment looks like, or what counts as sufficient bias mitigation) to later determination.[102] The AAA's text itself acknowledges this trade-off, using open-ended language ("to the extent possible", firms should consult external experts, etc.) that could dilute its force if not bolstered by strong FTC rules.[103] Overall, the **regulatory approaches differ in strictness and detail**: the EU opts to set substantive *rules of the road* for AI (leveraging its market power to raise global standards), while the AAA emphasises a *governance process* that integrates AI oversight into organisations' operations without prescribing one-size-fits-all solutions.

4.3 Redress Mechanisms

A crucial dimension of effectiveness is how each framework empowers oversight and remedies for wrongs. Here, the **EU regimes provide individuals with direct rights and avenues for redress**, whereas the AAA primarily relies on organisational compliance and government enforcement rather than individual claims. Under the **GDPR**, individuals (data subjects) enjoy an array of enforceable rights. If an AI-based decision infringes their rights or privacy, a person can lodge a **complaint with an independent Data Protection Authority (DPA)** in any EU member state (GDPR Art. 77) and is guaranteed an inquiry and a response.[104] Individuals also have the right to an **effective judicial remedy** against data controllers or processors (Art. 79) and can sue for **compensation** if they suffered material or non-material damage from a GDPR violation. In practice, this means a person negatively affected by an automated decision, for example, denial of credit by an algorithm, can appeal to their national DPA or court to investigate whether the decision violated data protection law (e.g. was there unlawful profiling or lack of transparency?). The GDPR's **one-stop-shop mechanism** further ensures cross-border complaints are handled in a coordinated fashion across the EU. Additionally, GDPR Articles 13–15 require that individuals be informed about **automated**

[102] Ibid 765-766.
[103] Ibid 766-767.
[104] Oskar Josef Gstrein and Andrej J Zwitter, 'Extraterritorial Application of the GDPR: Promoting European Values or Power?' (2021) 10 *Internet Policy Review* s 2.3 https://policyreview.info/articles/analysis/extraterritorial-application-gdpr-promoting-european-values-or-power

decision-making, including meaningful information about the logic involved, as part of transparency rights. This has been interpreted as a limited "right to explanation" of algorithmic decisions, enabling individuals to seek and receive an explanation of how a decision about them was made. [105]

The **AI Act** builds upon this rights-based enforcement model. While the Act will chiefly be enforced by designated supervisory authorities in each Member State, it notably introduces at least one explicit individual right: **Article 86 of the AI Act grants any person subject to a high-risk AI system the right to an explanation regarding its use**.[106] In other words, if an organization deploys a high-risk AI (for example, an AI system used in hiring or creditworthiness assessment), affected individuals can request an explanation of how that AI system's output influenced the outcome.[107] This provision echoes the GDPR's protections for algorithmic transparency, reinforcing individuals' ability to understand and challenge automated decisions. For enforcement, the AI Act establishes a European Artificial Intelligence Board for coordination, but on the ground, it relies on **national competent authorities** (which could be existing bodies or new ones) to supervise compliance and handle complaints. There is debate as noted by commentators about **whether Data Protection Authorities should take on AI Act enforcement** given their experience with AI-related cases under GDPR.[108] Regardless of which agency is tasked, EU individuals will be able to alert authorities if an AI system violates the Act (for example, if an AI provider failed to undergo conformity assessment or if a prohibited practice is suspected). Remedies can include orders to stop processing, recalls of non-compliant AI systems, and administrative fines. The **sanctions** in the AI Act are tiered and severe: up to **€35 million or 7% of global turnover** for the most serious violations (like deploying banned AI), and lower tiers (3% or 1% of turnover) for other breaches.[109]

---

[105] Joanna Mazur, Claudio Novelli and Zuzanna Choińska, 'Should Data Protection Authorities Enforce the AI Act? Lessons from EU-wide Enforcement Data' (12 June 2025) SSRN 6–10 https://ssrn.com/abstract=5290513 or http://dx.doi.org/10.2139/ssrn.5290513
[106] Ibid 15-16.
[107] Ibid 16.
[108] Ibid 16-18.
[109] AI Act, art 99.

These mirrors or exceed the GDPR's fine levels, indicating the EU's intent to back AI rules with real teeth.

In contrast, the **Algorithmic Accountability Act's enforcement and redress model is agency-centric.** The AAA does **not create new private rights of action** for individuals harmed by automated decisions, nor does it set up a dedicated complaint mechanism for consumers to challenge algorithmic outcomes. Instead, the Act would be enforced principally by the **Federal Trade Commission**, using its existing powers to police unfair or deceptive practices.[110] A violation of the AAA (such as failure to perform required impact assessments or address identified risks) is deemed an **"unfair or deceptive act or practice"** under the FTC Act, enabling the FTC to investigate and bring enforcement actions.[111] The FTC could issue orders, seek injunctions, and impose civil penalties on companies that do not comply. Notably, the AAA grants the FTC power to **impose penalties consistent with its usual enforcement (including monetary fines)** and to craft additional rules as needed.[112] Additionally, state attorneys general can enforce the Act on behalf of state residents (bringing civil actions in federal court if a company's violation harms people in their state).[113] This enforcement design is analogous to many U.S. consumer protection laws where government entities act as the enforcers; individuals themselves would rely on regulators to take action. Under the AAA, if a consumer believes an algorithmic decision was discriminatory or unlawful, they could complain to the FTC or their state AG, but **they would not have a statutory right to directly demand an explanation or sue the company** under this Act. Furthermore, the AAA does not mandate that companies **provide individuals with recourse or human review** of AI decisions. This is a notable divergence from GDPR's Article 22, which guarantees individuals the right *not* to be subject to purely automated decisions with significant effects without certain protections, including the possibility of human intervention. The AAA instead focuses on **organizational accountability**: as long as the company has done its due diligence (impact

[110] Algorithmic Accountability Act of 2025, S 2164, 119th Cong (2025) (introduced 20 June 2025)
[111] Ibid.
[112] Ibid.
[113] Ibid.

assessments, bias testing, etc.) and there is no violation of an existing anti-discrimination law or consumer protection law, the algorithmic decision can stand. In summary, the **EU frameworks empower individuals as enforcers of their own rights (backed by regulators and courts)**, whereas the **AAA places enforcement in the hands of regulatory authorities (the FTC and state officials), with compliance duties falling on organizations rather than rights directly exercisable by individuals**. The practical implication is that European citizens have more formal channels to challenge AI-driven decisions, while U.S. consumers under the AAA would be **shielded indirectly** (through oversight of companies) rather than through explicit personal rights vis-à-vis the algorithm.

4.4 Cross-Border Issues

In our increasingly global AI ecosystem, providers often operate across jurisdictions – meaning they must navigate **inconsistent legal definitions, risk categorizations, and procedures** in the EU versus U.S. regimes. A company like a large AI-driven platform or cloud AI service may be subject simultaneously to European requirements (if it has users in Europe or offers products there) and, if operating in the U.S., to emerging American rules like the AAA. This creates a complex **compliance landscape** where global AI providers must reconcile different regulatory vocabularies and expectations.

One key difference lies in the very **definition of AI systems vs. automated decision systems.** The **EU AI Act** defines an *"AI system"* broadly (covering software using machine learning, logic-based approaches, statistical methods, etc.) and then scopes its rules to systems deemed high-risk by sector/application.[114] The **AAA**, on the other hand, avoids the fraught question of defining "artificial intelligence" and instead targets *"automated decision systems" (ADS)* used in *"critical decision*

[114] AI Act, art 3.

processes." This terminology is technology-agnostic and **focuses on the function and impact** of a system rather than the fact that it is AI per se. [115] As scholars have noted, the AAA's focus on ADS for critical decisions (like those affecting access to credit, employment, education, etc.) captures essentially the same high-impact use cases as the AI Act's high-risk category, but **without requiring a judgment on what counts as an AI system**.[116] For a multinational provider, this means the **trigger for compliance** might differ: in the EU, the company must first determine if its product includes an AI component as defined by the Act and if it falls into a listed high-risk area (Annex III of the AI Act), whereas in the U.S., the focus will be on whether the system is being used to make significant decisions about consumers (regardless of whether it uses advanced AI or simpler automation). In practice, most systems covered by the AAA (loan approval algorithms, hiring filters, insurance pricing models, etc.) would also be considered high-risk AI under EU law but there could be edge cases. For example, an expert system using hand-coded rules for credit underwriting might be subject to AAA (because it's an automated decision system with a material effect) yet **not technically meet some definitions of AI** (if one takes AI to mean machine learning) the EU Act's broad definition likely still captures it, but any mismatch in definitions could cause uncertainty. Providers will likely err on the side of treating any automated decision tool as potentially in scope in both regimes, but they must be aware of these conceptual differences.

Another divergence is in **risk classification versus case-by-case risk assessment.** The **EU AI Act imposes a *taxonomy* of risk levels**, with predefined categories and compliance routes: e.g. biometric identification systems are designated as high-risk and must follow strict requirements, real-time facial recognition in public is banned as unacceptable, etc. This **categorical approach** means a provider must first see if their system fits a listed category (like AI for recruitment is explicitly high-risk) and then follow the prescribed steps (conformity assessment, CE marking, etc.) before deployment. The **AAA**, in contrast, does not assign static risk labels like "high" or "low" to particular applications. Rather, it uses a **contextual, firm-driven risk analysis**: companies must

[115] Mökander, Juneja, Watson and Floridi (n 99) 756-758.
[116] Ibid 757-758.

assess and document the risks of their automated systems and mitigate those deemed likely to harm individuals (for example, disparate impact or privacy intrusions). There is no equivalent list of "unacceptable AI practices" under the AAA that must be banned. A practice unacceptable in Europe (say, social scoring of individuals) would only be illegal in the U.S. if it violates some other law or if the FTC deems it unfair. This places a burden on global AI providers to **adapt their compliance procedures**: in the EU they will follow a **checklist of requirements** for each high-risk system (data quality, transparency, human oversight, etc. as explicitly detailed in the AI Act), whereas in the U.S. they will need to craft an **Algorithmic Impact Assessment report** addressing somewhat analogous topics (accuracy, fairness, privacy, security) but *in their own terms*, subject to eventual review by the FTC. Providers may find it efficient to develop a **unified internal AI governance process** that meets the stricter elements of both regimes, for instance, performing a rigorous assessment (as per AAA) and also ensuring all the specific EU technical documentation is generated. However, conflicts could arise: an AI system might **pass U.S. muster** (no evident bias found in the company's assessment) yet **fall short of EU requirements** (perhaps it lacks a required level of human oversight or fails to meet the AI Act's quality criteria). Global companies will likely have to implement the more stringent standard in each area to be safe, effectively adopting a **"highest common denominator"** approach to compliance.

Procedural obligations and timelines also differ cross-border. The AI Act demands **pre-market conformity assessment** for many systems, possibly involving external auditors or notified bodies, before the system can be sold in Europe.[117] This could slow down deployment as companies gather certifications. In the U.S., the AAA does not require pre-approval; a company can deploy and then later submit an assessment (the bill envisions that summary reports of impact assessments may be filed with the FTC within a certain period). Thus, a start-up launching an AI tool internationally might face a *delay in the EU* to complete compliance steps, while in the U.S., it could launch and then adjust on the fly. Over time, one might see products **initially released in the U.S. market and**

[117] Ibid 759-760.

only later in the EU after meeting the AI Act's formalities, echoing patterns seen with some medical devices and privacy-sensitive products under GDPR. Additionally, the **legal definitions of protected risks and harms vary**: the AI Act has a broad fundamental-rights perspective (e.g. it considers an AI system's potential harm to EU values and rights), whereas the AAA explicitly enumerates **protected classes and traits** (race, colour, sex, religion, etc.) in requiring evaluation of differential impacts.[118] A global AI provider must ensure their system's fairness across these dimensions; the concept of "bias" might be framed slightly differently in EU discourse (which might talk more about discrimination or dignity) versus U.S. discourse (which often frames it in terms of civil rights law compliance). **Global providers also face duplicative procedural obligations;** for instance, both the AI Act and AAA might require some form of **impact/risk assessment**, but the format and disclosure differ. Companies will likely compile one extensive assessment internally, then **extract subsets** to satisfy each law. The AI Act may require sharing documentation with EU regulators or users (on request), while the AAA may require submitting a summary to the FTC and keeping detailed records available for inspection. Navigating these will demand robust internal governance: many firms are already creating **AI compliance teams** to map the requirements of each regime and ensure a coherent strategy.

Finally, cross-border AI providers must reconcile **differences in oversight and liability**. A European deployment exposes the company to potential DPA investigations or European court orders, whereas a U.S. deployment (if AAA is enacted) could trigger an FTC investigation or state attorney general action. The threshold for intervention might differ. European authorities might proactively audit high-risk AI providers even absent a breach (since registration/conformity info is filed), whereas the FTC often acts after a problem comes to light (e.g. a scandal or complaint). This means **risk appetite and compliance culture** may need to be uniformly high: a lapse in one jurisdiction could quickly become a global public relations issue. Inconsistent legal definitions and standards may also incentivise **international regulatory cooperation**. Indeed, there are ongoing

[118] Ibid 761-762.

transatlantic dialogues (through the EU–US Trade and Technology Council and other fora) seeking to align AI governance approaches to reduce friction. Until greater alignment is achieved, however, global AI actors must treat the EU and US frameworks as **distinct compliance tracks**, carefully tracking where one mandates something the other does not. For example, an AI developer might decide to implement the **strictest transparency measures globally**, such as always disclosing AI use and providing an explanation on request, to satisfy the EU, even though U.S. law might not require proactive disclosure. Conversely, companies may lobby for U.S. regulators to accept compliance with the EU AI Act as a showing of good faith or due diligence under the AAA. In sum, the current cross-border reality is that AI providers face a patchwork of definitions (what is "AI" or "ADS"), **different categorisations of risk, and different procedural hoops**. This demands a **synchronised compliance strategy** to avoid legal gaps: many firms will end up implementing **dual (or multiple) compliance programs** for their AI systems, incurring higher costs but hopefully ensuring that their systems meet all applicable standards of safety, fairness and accountability worldwide.

4.5 Comparison between six key dimensions: transparency and explainability; risk classification and mitigation; data subject rights and human oversight; enforcement and sanctions; and innovation/regulatory flexibility

In evaluating the effectiveness of the GDPR, EU AI Act, and AAA, it is helpful to compare how each framework addresses key governance dimensions. Table 4.1 below provides a summary across six core dimensions, followed by a concluding discussion of the frameworks' relative strengths, gaps, and likely evolution:

Table 4.1 – Comparison of GDPR, EU AI Act, and US AAA Across Key Governance Dimensions

Dimension	GDPR (EU, 2018)	EU AI Act (2024)	US Algorithmic Accountability Act (proposed)
Transparency & Explainability	Emphasizes transparency in personal data processing (Arts. 12–14 GDPR); individuals must be informed of automated decision-making logic to a meaningful extent. Debate over a "right to explanation," but GDPR ensures data subjects can request and receive rationale for significant automated decisions.	Requires transparency for certain AI systems: e.g. AI interacting with humans must disclose it is AI, AI-generated content must be labeled. High-risk AI providers must ensure **explainability** such that outputs can be understood by users. Article 86 gives individuals a right to an explanation of outcomes from high-risk AI, strengthening explainability.	Primarily internal transparency. Companies must **assess and document** their AI systems' transparency and explainability as part of impact assessments. No affirmative right for consumers to be informed when AI is used or to receive an explanation of a decision. Any public transparency (e.g. summary reports to the FTC, or notices) will depend on FTC rules and existing sectoral laws.
Risk Classification & Mitigation	Does not categorize AI systems by risk but uses a general risk-based approach in data protection (e.g. DPIAs for high-risk processing). Implicitly addresses algorithmic risks via principles (data minimization, fairness) and case-by-case oversight (Art. 22 limits high-impact automated decisions). Mitigation of risks is ensured through compliance with broad principles and corrective measures by DPAs.	Employs an explicit **risk-based model**: AI systems classed as *unacceptable risk* (banned), *high-risk* (allowed with strict safeguards), *limited risk* (some transparency), or *minimal risk* (no requirements). **Mandates risk mitigation plans** for high-risk AI (ongoing risk management, quality control, human oversight, etc.). Built-in process to update risk classifications as technology evolves, via delegated acts.	No fixed risk tiers; **focuses on "critical decision" use-cases** (significant impacts on consumers' lives) rather than labeling levels of risk. All covered systems effectively treated as high impact, but none banned outright by this Act. Mitigation is through required **Algorithmic Impact Assessments** and follow-up actions: firms must identify biases or hazards and "reasonably" address them. The law relies on companies to calibrate and reduce risks, subject to later review.

Dimension	GDPR (EU, 2018)	EU AI Act (2024)	US Algorithmic Accountability Act (proposed)
Data Subject Rights & Human Oversight	Strong individual rights: access, rectification, erasure, objection, and *not to be subject to certain automated decisions* without safeguards (Art. 22). Individuals can demand human intervention in automated decisions and express their viewpoint or contest the outcome. GDPR effectively requires human oversight for high-stakes AI by giving persons the right to have a human review automated decisions.	Focuses on systemic oversight rather than individual rights (aside from the explanation right). **Mandates human oversight measures for high-risk AI** – providers must design systems to be overseen by humans or provide for human control. Deployers of high-risk AI are expected to ensure appropriate human monitoring of the system in operation. No new individual "opt-out" or decision-specific appeal rights in the Act (those remain under GDPR for personal data scenarios), but the required human oversight is analogous to ensuring a "human in the loop" for critical AI decisions.	No explicit individual rights vis-à-vis automated decisions. The AAA does **not guarantee consumers the right to human review or to opt out** of AI-driven decisions. It also does not require companies to obtain consent for or offer alternatives to automated decisions. **Human oversight** is not mandated, though the Act's title ("Augmented Critical Decision Processes") implies many covered systems will involve human+AI hybrid decision processes. Any human-in-the-loop practices are left to the organization's discretion or to sector-specific laws.

Dimension	GDPR (EU, 2018)	EU AI Act (2024)	US Algorithmic Accountability Act (proposed)
Enforcement & Sanctions	Enforcement by independent DPAs in each EU state, cooperating via the EDPB. DPAs can investigate, issue orders (e.g. ban processing), and impose fines up to **€20 million or 4% of global annual turnover** for violations. Individuals have the right to lodge complaints with DPAs and sue in court; GDPR fines have been in the hundreds of millions for big tech[6]. Robust public enforcement coupled with private rights (complaints, collective actions possible via Article 80).	Enforcement primarily by national **Market Surveillance Authorities** (could be sector regulators or new AI authorities), coordinated by a European AI Board. Severe penalties: up to **€35 million or 7% of global turnover** for the worst breaches (e.g. unlawful use of banned AI). Lesser offenses carry up to 3% or 1% of turnover. Individuals can file complaints about AI systems to authorities (exact mechanisms to be set by each country). No direct damages claim under the Act, but harmed parties might sue under general tort or product liability regimes as complemented by the Act.	Enforcement by the **FTC** using its existing powers (treats violations as unfair/deceptive acts). The FTC can seek injunctions, consent orders, and civil penalties (which can reach tens of thousands of dollars per violation, and potentially more under FTC's penalty authority for rule violations). State Attorneys General can also enforce the law in court. No automatic fines like GDPR's percentages, but enforcement actions could result in significant remedial orders or monetary penalties (e.g. FTC algorithms cases to date have led to algorithms being deleted or products banned). Importantly, **no private lawsuits under AAA** – individuals must rely on regulators to act.

Dimension	GDPR (EU, 2018)	EU AI Act (2024)	US Algorithmic Accountability Act (proposed)
Innovation & Regulatory Flexibility	Proponents credit GDPR with instilling **trust and quality** in data practices, which can spur innovation in privacy-friendly AI. Critics argue GDPR's strict rules (and heavy compliance costs) burden AI development, especially for data-intensive innovation and SMEs. The law is technology-neutral but relatively inflexible once in force – changes require legislative amendment. However, GDPR has driven convergence toward higher data protection standards globally (the "Brussels effect"), potentially leveling the playing field and encouraging **compliance-driven innovation** (firms innovate within the guardrails).	The AI Act takes a **proportionate approach** – lighter or no requirements on low-risk AI – aiming not to over-regulate benign innovations. It includes measures to support innovation, such as **sandboxes** for AI developers and reduced compliance fees for SMEs (Article 55). Still, high-risk AI developers will face substantial compliance overhead (which some fear could deter startups or drive AI research out of the EU). The Act is more easily updated than a statute via delegated acts for technical standards, providing some flexibility. Overall seen as Europe's attempt to set **global AI standards** early, potentially giving compliant companies a competitive edge in trustworthiness. But there is a trade-off: stringent requirements might slow down release of new AI products in the EU while the rest of the world iterates faster.	The AAA's philosophy is to **minimize regulatory burden while addressing the worst risks**. It explicitly limits scope to large companies, exempting startups and small businesses from direct obligations. By focusing on process (impact assessments) rather than dictating design, it offers companies flexibility in how they mitigate risks – ideally fostering **innovation with accountability**. The Act's requirements could institutionalize ethical AI practices without prescribing technical solutions, allowing firms to continually experiment. Also, because much is left to FTC rulemaking, the regulatory regime can adapt over time and respond to industry input. However, the AAA is much less comprehensive than the EU approach – critics suggest it may **lack teeth or leave gaps** (e.g. AI used by small firms or local governments falls outside its scope). In terms of innovation, the lighter touch may encourage AI rollouts, but without uniform standards some risky systems might proliferate unchecked. The AAA's success in balancing innovation and protection will heavily depend on robust FTC enforcement and updates.

5. Critical Perspectives and Gaps

In this chapter, we examine key critiques and gaps in the emerging AI regulation landscape. A crucial decision is how to structure this discussion. One approach is **thematically**, highlighting cross-cutting issues (like ambiguity, overbreadth, transparency vs. trade secrets) and comparing how each regime handles them. This can illuminate common challenges and avoid repetition. The other approach is an **integrated narrative by regime**, analysing EU and US frameworks in turn. Given that earlier chapters likely treated each regime separately, a thematic structure here could offer a fresh comparative lens; however, it may require careful transitions to ensure coherence. Below, we outline the major critical perspectives organised by topic, but with comparisons across GDPR, the EU AI Act, and US proposals, followed by a brief international outlook.

5.1 GDPR's Limitations and the "Right to Explanation"

Scholars have long debated the scope of GDPR Article 22's protection against automated decisions. **Ambiguity and Low Invocation:** Article 22's wording ("right **not** to be subject to a decision based **solely** on automated processing that produces legal or similarly significant effects") leaves room for interpretation. Critics note that the **"solely"** qualifier is a loophole that even trivial human involvement could arguably bypass the prohibition. Wachter et al. famously argued that a controller could skirt Article 22 by inserting a nominal human "rubber stamp," rendering the decision not *solely* automated. [119]Although the European Data Protection Board (EDPB) warned against such "fabricated" human involvement, the very need for this guidance underscores Article 22's ambiguity. Indeed, a strict reading suggests Article 22 **"is very rarely applicable,"** as many automated systems involve at least minimal human input in practice. Consequently, enforcement

[119] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 76, 81-82 https://ssrn.com/abstract=2903469 or http://dx.doi.org/10.2139/ssrn.2903469

has been limited until recently there were virtually no landmark cases interpreting Article 22, and it has been invoked infrequently in legal challenges.

Explainability Guidance Gaps: GDPR technically stops short of an explicit "right to an explanation" in the operative text, mentioning only a right to receive "meaningful information about the logic involved" (Articles 13–15) when automated decisions are made. For years, what counts as **"meaningful"** information remained vague, providing limited practical guidance to organisations on how to explain AI decisions.[120] This vagueness has been a point of critique, as it left both data subjects and controllers uncertain about how much detail algorithms must reveal. Only in 2025 did the CJEU finally clarify the standard: in a credit scoring case it held that controllers must give **"relevant information, in a concise, transparent, intelligible and easily accessible form, on the procedure and principles actually applied"** by the algorithm.[121] The CJEU emphasised that the explanation should include the **importance and envisaged consequences** of the processing, possibly with tangible examples. This ruling effectively confirmed a de facto *right to explanation* under GDPR. It also stressed a balance too little detail undermines transparency, but too much technical complexity overwhelms the data subject.[122] The late arrival of this clarity highlights the gap: for several years, firms struggled with how to provide explanations, and data subjects had patchy rights, due to scant official guidance on Article 22's requirements.

Transparency vs. Trade Secrets: Another critical gap has been the conflict between transparency obligations and the protection of trade secrets or IP in algorithms. Companies often resist disclosing how their AI systems work, citing proprietary models. GDPR's Recital 63 acknowledges that access rights should not adversely affect others' rights such as trade secrets. In practice, this led some controllers to refuse sharing meaningful algorithmic details. However, a February 2025 CJEU decision (Dun & Bradstreet case) pushed back on blanket trade-secret defenses. The Court held that **controllers cannot simply invoke trade secrets to deny information**; instead, an individualized

[120] Case C-203/22 *Dun & Bradstreet Austria* EU:C:2025:152, Judgment of the Court (27 February 2025).
[121] Ibid.
[122] Ibid.

balancing is required.[123] National courts or regulators can even compel companies to confidentially **disclose the secret algorithmic details to them**, so they can judge what must be revealed to the individual. This is a significant development because it chips away at the trade-secret shield – but it also raises practical challenges (e.g. most data protection authorities lack clear powers or expertise to evaluate trade secrets).[124] The underlying tension remains: GDPR mandates algorithmic transparency to an extent, while companies fear oversharing sensitive IP. Critics point out that until this recent case, there was little formal resolution of this tension, potentially discouraging individuals from exercising access rights and allowing controllers to err on the side of secrecy. Even now, businesses voice concern that forced disclosure (even if filtered through courts) could undermine incentives to innovate. In summary, GDPR's lofty principles left a gap between **transparency and confidentiality**, only partially addressed by evolving case law.

5.2 EU AI Act Uncertainties and Critiques

The EU's draft **Artificial Intelligence Act (AI Act)** has drawn a mix of praise and concern. As a pioneering comprehensive AI law, it faces **growing pains in its scope and implementation**:

- **Overbroad or Rigid Risk Classification:** The AI Act employs a risk-based "pyramid" that **categorises AI systems into four tiers** (unacceptable, high, limited, minimal risk) with corresponding rules. While this approach is intuitively appealing, stakeholders worry about **over-categorisation** – i.e. that the definitions of high-risk AI are too sweeping. For example, the Act would label many insurance industry algorithms as *"high-risk"* simply because they are used in life or health insurance underwriting. Insurers argued this is **"too far-reaching"**, capturing even routine scoring tools that *"are not likely to cause serious…risks"*. More generally, the high-risk category triggers onerous requirements (data governance, documentation, human oversight, etc.), and **critics fear it will encompass AI**

[123] Ibid.
[124] Ibid.

systems that pose only minimal risk, thus **imposing undue compliance burdens**. If innocuous or low-impact AI gets over-regulated due to broad definitions, innovation could be stifled and companies might face **"undue burden…not technically feasible"** to implement. The drafters have tried to fine-tune the risk criteria (e.g. excluding truly *"minimal"* software), but **uncertainty persists** about borderline cases (what exactly is a "similarly significant" risk to fundamental rights? how will *general-purpose AI* like ChatGPT be handled?). This over-categorization critique points to a gap in **predictability**: businesses may struggle to know if their system will be deemed high-risk, and regulators might be flooded with borderline cases, at least until further guidance or precedent emerges.

- **Regulatory Capture and the Voluntary Code of Practice:** A very recent development is the EU's *Voluntary Code of Practice on AI* (especially for general-purpose AI systems) introduced ahead of the AI Act's full enforcement. The Commission invited AI providers to **sign up to a Code of Practice for Generative AI**, promising those who do an easier compliance path ("legal certainty") once the Act kicks in. While this **public-private cooperation** can accelerate adoption of best practices, observers worry about **industry influence** dominating the process. Indeed, within days of the Code's release, major tech CEOs openly lobbied to **water down the AI Act itself** as "toxic" to innovation. **Critics warn of a risk of regulatory capture**: if regulators rely too heavily on industry-drafted standards, the resulting rules might be **"softened to fit incumbents' business models,"** at the expense of public interest protections. The voluntary Code could become a de facto certification that big players obtain, possibly even an excuse for regulators to give them a lighter touch while smaller firms struggle to keep up. Ethicists note that if **"economic power translates into regulatory bargaining power,"** the **legitimacy of the AI Act's framework could be undermined**. The Commission must therefore guard against the Code evolving into an industry wish-list. Ensuring **civil society participation** and independent oversight in the Code's governance has been suggested to counteract this risk. In short, the

AI Act's implementation via a voluntary scheme presents a gap between **policy and practice**: it's a novel attempt to include industry in rulemaking, but it raises **"unusually stark"** ethical questions about when it is appropriate for regulators to let those regulated *"help write – or rewrite – the rules"*. The outcome is uncertain: it could either enhance compliance or entrench the power of a few large AI providers.

Beyond these points, it's worth noting the **political feasibility** of the AI Act was a concern but now largely resolved as of 2025, the Act is in final negotiations (or newly enacted). However, **implementation challenges** (ensuring consistent enforcement across 27 Member States, developing the capacity for conformity assessments, etc.) remain a gap often cited by commentators. Some worry about a **"Brussels effect"** imposing the EU's stringent rules globally, while others fear uneven readiness within the EU but those debates go beyond the scope here. The critical perspectives above suggest the EU framework, while bold, must avoid **over-correcting** (over-regulating benign tech) and **capture** (letting the regulated set the rules).

5.3 U.S. Algorithmic Accountability Act (AAA): Gaps and Feasibility

In the United States, the **Algorithmic Accountability Act** in its various introduced versions (2019, 2022, and the latest 2025 bill) has been a central proposal to federally regulate AI. It too faces critiques about **coverage gaps** and uncertain prospects:

- **Limited Scope – Exempting SMEs and Others:** A common criticism is that the AAA only targets **large companies**, leaving a **regulatory gap for small and medium-sized enterprises (SMEs)**. For instance, the 2019 draft applied only to businesses with over $50 million annual revenue or holding data on >1 million consumers. The rationale was to avoid over-burdening startups and small firms. But many argue this carve-out is misguided: *"If a certain decision can be harmful…, the size of the company…does not seem relevant"*. A biased algorithm used by a small firm in hiring or lending can harm individuals just as much

as one used by Big Tech. Joshua New noted it *"makes little sense to exempt the majority of companies from compliance"* when serious AI risks can emerge anywhere. Thus, **many high-impact AI applications (especially those developed by smaller vendors or used by mid-size organizations) would escape oversight** under the AAA. This gap is especially salient given the startup-driven nature of AI innovation critics fear a two-tier system where big firms face audits and impact assessments, but "small" players (potentially up to hundreds of millions in revenue) fly under the radar. The latest 2025 Senate bill (S.2164) does expand coverage somewhat (lowering some thresholds to $5M revenue for certain contexts), but it still excludes truly small startups and any business outside FTC jurisdiction (e.g. most nonprofits and some financial institutions). Another gap is that **government use of AI isn't directly covered** by the AAA (which focuses on companies) algorithmic bias in public-sector tools (policing, benefits decisions, etc.) would not be addressed, an omission civil rights groups have noted.

- **Enforcement and Trade-off Questions:** Even for those entities that would be covered, the AAA's requirements (algorithmic impact assessments, audits, transparency reports) raise feasibility concerns. Business advocates warn of heavy compliance costs, especially if every new model or update requires an assessment. There is also debate over **publishing the impact assessments**: the AAA does *not* currently require making them public, to protect trade secrets, which some see as a missed opportunity for transparency. On the other hand, making them public could discourage frank internal analysis. These design choices reflect a tension between **accountability and corporate confidentiality** a theme similar to the GDPR trade secret issue, but in the context of audit reports. Moreover, absent strong enforcement mechanisms, there's scepticism about how effective the AAA would be even if passed. The FTC would need significant resources and expertise to oversee algorithmic assessments nationwide – and without dedicated funding, the law could become a paper

tiger. These concerns highlight gaps in **operationalizing** the lofty principles of accountability.

- **Political Feasibility – Uncertain Enactment:** Perhaps the largest "gap" is that, as of 2025, the AAA is **still not law**. Earlier iterations have stalled in Congress. The regulatory momentum in the US has lagged behind the EU in part due to partisan splits and industry pushback. In fact, whenever a comprehensive bill is introduced, counter-proposals or disagreements arise (for example, a competing bill in 2020 omitted algorithm audit provisions, reflecting divergent views in Congress). The AAA of 2022 died in committee, and while reintroduced in 2023/2025, its passage is **not guaranteed**. Lawmakers and commentators have acknowledged it is "unlikely to pass in the near term" given the current political climate. This uncertainty has real consequences: it leaves the U.S. with *de facto* no federal AI-specific regulation, meaning the very risks the AAA seeks to mitigate remain governed only by patchy general laws (like anti-discrimination statutes or the FTC's broad consumer protection authority). The **prospect of the AAA not being enacted at all is very real**, raising questions about what fallback mechanisms might address algorithmic harms (hence the White House's non-binding *AI Bill of Rights* as one attempt, and reliance on state laws as discussed next). In summary, the AAA is a well-intentioned framework that, according to critics, **covers too few actors to truly protect the public, and still faces an uphill battle to become actual policy**. Until those gaps are addressed, federal oversight of AI in the US remains more aspiration than reality.

5.4 The Patchwork Problem: State Laws and Sectoral Rules in the US

In the absence of comprehensive federal AI regulation, a **patchwork of state laws and sector-specific rules** is rapidly emerging in the United States. While these can plug some gaps, they also introduce new challenges:

- **Divergent State AI Laws:** Several states have forged ahead with their own AI or algorithmic accountability statutes. **Colorado** became the first in 2024 with a law addressing AI-driven decisions in areas like employment, housing, credit, and healthcare. It **mandates impact assessments, bias mitigation, transparency, and human alternatives** for "high-risk" AI systems, for both developers and deployers, effective 2026. On one hand, this is a milestone for AI accountability. On the other, even Colorado's governor signed the bill with "visible hesitation," warning it creates a **"complex compliance regime"** that might **"tamper with innovation and deter competition"**, and explicitly urged Congress to step in with a uniform approach. Attempts to narrow the Colorado law's scope or delay it (and to exempt small businesses) were unsuccessful. Meanwhile, other states are crafting *their* versions: e.g. **Virginia** considered a similar bill but it was vetoed as overly rigid; **Texas** debated an EU-style AI Act copycat; **California** has introduced an **Automated Decision Systems bill** imposing audits and opting-out rights; **New York** (state) has looked at rules for AI in critical decisions, and **Illinois** has a new law barring AI in video interviews that cause bias, etc. The net result is a **fragmented regulatory map**, where compliance requirements differ by jurisdiction. Each state defines "AI" and "high-risk" with its own nuances. A system might be deemed high-risk in one state but not in another; one state might demand an annual bias audit, while another requires specific notices to consumers.

- **Inconsistent Sectoral Regulations:** Apart from these general AI laws, sector-specific regulations are proliferating. For instance, in hiring and employment: **New York City's** Local Law 144 (effective July 2023) forbids employers from using Automated Employment Decision Tools (like AI resume screeners) unless an independent bias audit is conducted annually and candidates are notified. **Illinois** and other locales have their own rules on AI in hiring or interviews. In insurance, regulators are examining or issuing guidance on algorithmic underwriting fairness. Federal agencies, too, have domain-specific initiatives (e.g. the EEOC's enforcement guidance on AI hiring tools under discrimination law, the

CFPB's focus on AI in credit). While these targeted efforts address real risks, together they create a **maze of overlapping obligations** for any company operating across state lines or sectors. A tech firm offering an AI tool nationally might have to comply with a dozen different standards – a far cry from the single set of rules under the EU AI Act.

- **Compliance Burden and Uncertainty:** Observers describe this as a **"chaotic patchwork"** that is **rapidly multiplying legal complexity**. Each new law adds a layer of paperwork (audits, disclosures, notices) which, collectively, can overwhelm especially smaller companies. Firms are hindered in their ability to **"innovate and compete"** when facing 50 different AI regimes. The lack of uniform definitions and requirements can also undermine the very purpose of regulation: if rules are too inconsistent or confusing, some organizations may inadvertently fall foul of them or even avoid certain jurisdictions altogether. Economically, this fragmentation **"disrupts economies of scale"** an AI product cannot be rolled out uniformly if, say, it's banned in one state and allowed in another and may **divert AI investment to 'friendlier' states or countries**. Ultimately, there is a broad consensus that a baseline federal law (like the AAA) would be preferable to this Balkanization. Without it, the U.S. could see a race where stricter states push heavy compliance costs while lenient states undercut with lax rules, neither of which is ideal. In the words of one policy analyst, *"the United States cannot afford a fractured market with 50 competing AI regimes."*[125] Yet, as noted, federal action is uncertain, so companies must navigate the patchwork for now.

In sum, the patchwork of state and sectoral measures fills some **regulatory gaps** (ensuring at least some oversight of AI where federal law is absent), but at the cost of **inconsistency and complexity**. This raises compliance risks for businesses and leaves individuals with unequal protections

[125] Kristian Stout, 'Federal Preemption and AI Regulation: A Law and Economics Case for Strategic Forbearance' (WLF Legal Pulse, 30 May 2025) https://www.wlf.org/2025/05/30/wlf-legal-pulse/federal-preemption-and-ai-regulation-a-law-and-economics-case-for-strategic-forbearance/

depending on where they live. It's essentially the opposite of the EU's harmonized approach and it illustrates a major gap in U.S. governance of AI: the lack of a unified strategy.

5.5 International Comparison and Other Gaps

No discussion of AI governance gaps would be complete without a brief look beyond the EU and US, as other jurisdictions are also grappling with these issues:

- **Canada:** Canada's effort to regulate AI federally through the **Artificial Intelligence and Data Act (AIDA)** has faced setbacks. *As of mid-2025, Canada has no approved AI-specific law* – AIDA was introduced as part of a bill in 2022 but stalled and had to be reintroduced after a recent election. The proposed AIDA took a risk-based approach (targeting "high-impact" AI systems in sectors like healthcare, employment, finance, etc.) similar to the EU model. However, it has been **criticized for gaps**, such as **no clear definition of "high impact"** (leaving it to regulators later) and lack of explicit human rights protections**. Observers noted that the initial draft** failed to include key stakeholders **and may not adequately protect marginalized groups. In short, Canada's regulatory push is in flux – the gap here is one of** uncertainty and delay**, meaning AI oversight currently relies on general data protection and sectoral laws. Canadian companies must watch both domestic developments and the need to comply with EU or US rules when exporting AI products.

- **United Kingdom:** The UK has deliberately taken a *"pro-innovation"* and light-touch stance instead of a single omnibus AI law. In 2023, the UK government released an AI White Paper advocating a **principles-based, sector-led approach**: regulators in various domains (health, finance, etc.) are tasked with applying broad principles like safety, transparency, fairness to AI within their remit, rather than new legislation. This flexible approach aims to avoid stifling innovation, but critics point out its **shortcomings**: the voluntary guidelines lack teeth, leading to **regulatory uncertainty** about what standards actually apply. There is also a

risk that without binding rules, **important safeguards could be delayed or diluted**, and the UK's divergence from the EU might create compliance headaches for companies operating in both. Essentially, the UK is testing a gap-filling strategy that leans on existing laws – time will tell if this patchwork (different from the US patchwork, as it's coordinated by principle) can match a unified law in effectiveness. The **"all eyes and no hands"** critique is sometimes leveled: lots of oversight bodies watching AI, but no single authority with strong enforcement, potentially leaving enforcement gaps.

- **China:** China has moved quickly on AI regulations, but with a very different focus. Its 2022 **Algorithmic Recommendation Regulation** and 2023 draft rules on generative AI put heavy emphasis on **content control, security, and social stability** – for example, requiring recommendation algorithms to promote approved content and generative AI to reflect core socialist values. These laws fill the gap on issues like disinformation and censorship (indeed, they are more stringent than Western laws in that area), but they **do not prioritize transparency or individual rights** in the way GDPR or the AI Act do. One could say China's approach leaves gaps in **privacy and freedom**, even as it closes gaps in government oversight and data sovereignty. Additionally, much of China's AI governance happens via administrative guidelines (which can be rapidly updated), raising concerns about consistency and due process. For international companies, China's rules represent yet another regime to navigate – often incompatible with Western values – illustrating the global compliance challenge.

- **Other Jurisdictions:** The EU's approach has influenced some neighbors e.g. Brazil and India are considering AI frameworks, and OECD countries broadly endorse ethical AI principles – but concrete laws remain rare. Some countries (like **Japan**) emphasize an agile governance approach, addressing issues as they arise rather than prescriptive rules, which may leave **gaps in preparedness** for future AI risks. Others focus on specific matters (e.g. Australia on AI in safety-critical systems). Overall, there is a patchwork at the international

level too, with **divergent regulatory philosophies**. This complicates the creation of any **global standards** and might allow regulatory arbitrage (AI developers relocating to jurisdictions with fewer rules).

International comparisons highlight that while the EU and (to a lesser extent) the US are forging ahead with AI regulations, much of the world is still catching up or taking a different path. This creates a **risk of inconsistent protections worldwide** for instance, a European's rights against an AI decision are clearly defined (if not always enforceable), whereas a person in a country without AI laws might have no specific recourse. It also poses challenges for companies trying to **comply across borders**, as they must reconcile very different legal expectations. These gaps suggest a need for at least *some* coordination perhaps via international standards or treaties to prevent AI governance from itself becoming fragmented.

6. Case Studies

This chapter examines the effectiveness of existing and proposed regulatory frameworks in governing AI-powered decision-making through three detailed case studies. We focus on: **(1) Clearview AI**, a company offering facial recognition technology built on a massive biometric database scraped from the internet; **(2)** SCHUFA**, Germany's leading credit scoring agency whose automated credit assessments have been scrutinised under Article 22 GDPR; and** (3) HireVue**, a provider of AI-driven hiring tools used to screen job candidates. These case studies span distinct domains: law enforcement surveillance, financial services, and employment yet all raise common issues about how to ensure** legal compliance**,** ethical principles **(fairness, transparency, non-discrimination, accountability),** and effective oversight for AI-driven decisions.

We evaluate each case under the EU's General Data Protection Regulation (GDPR) and the forthcoming EU Artificial Intelligence Act (AI Act), as well as under the lens of the proposed U.S. **Algorithmic Accountability Act (AAA),** a legislative initiative aiming to impose algorithmic transparency and fairness obligations. For each case, we review compliance with the law (or lack thereof), the ethical implications of the AI system's deployment, the role played by regulators (such as Data Protection Authorities in the EU and the U.S. Federal Trade Commission), and any enforcement outcomes. Finally, we distil **lessons learned** from each case and discuss implications for future regulatory strategies to govern AI in a manner that upholds fundamental rights and societal values.

6.1 Clearview AI: Facial Recognition and Biometric Surveillance

Case Overview: Clearview AI is a U.S.-based company that built a controversial facial recognition database by scraping over 3 billion images of faces from social media and other public websites. The company's software allows users (including law enforcement agencies) to upload a photo of an individual and algorithmically match it against this database, returning any identified matches along with metadata. Clearview marketed this tool to police as a way to identify unknown individuals (suspects, persons of interest, or even incidentally photographed people) by linking their faces to online profiles. Essentially, Clearview turned publicly available personal photos into a **gigantic biometric surveillance system**, largely without the knowledge or consent of the individuals affected.[126]

Legal Compliance under GDPR: Clearview's activities squarely conflict with core GDPR provisions. European Data Protection Authorities (DPAs) in multiple jurisdictions have found Clearview's data processing unlawful on numerous grounds: **lack of a legal basis**, unlawful

[126] van Haperen (n 67) 10–11.

processing of sensitive data, opacity, and disregard for data subject rights. Clearview scraped personal photos en masse, but **did not obtain consent** from individuals (indeed, data subjects were unaware of the processing).[127] Nor could Clearview credibly rely on an alternative lawful basis like "legitimate interests", given the severe intrusion into privacy and lack of reasonable expectations by individuals that their online photos would be harvested for facial recognition.[128] In GDPR terms, the images may have been publicly accessible, but they were *not* "manifestly made public" in a manner that would waive privacy rights, and DPAs emphasized that public personal data are still fully protected by the GDPR. Clearview also processed biometric data (facial vectors) at scale, which are **special category data** under GDPR (Art. 9); none of the narrow exceptions (explicit consent, vital interest, etc.) applied.[129]

Clearview further violated **GDPR principles of transparency and purpose limitation**. It **provided no notice or information** to data subjects (breaching Articles 12–14) and was found to have **impeded data subject access requests** – for example, by failing to respond adequately or at all, and by arbitrarily limiting the scope of responses (only providing data from the preceding year, etc.). Such conduct flouts the right of access (Art. 15) and the right to erasure (Art. 17), both of which were invoked by individuals in Europe to no avail. Moreover, Clearview had no established representative in the EU as required by Article 27 for foreign controllers, another compliance failure noted by regulators. In sum, under the GDPR's regime, Clearview's model, which collects Europeans' photos for facial recognition without consent or other valid legal ground, and processes highly sensitive biometric data in a secretive way, is fundamentally unlawful.[130]

DPAs across Europe reached that exact conclusion. In late 2021 and 2022, authorities in **France, Italy, Greece, and the UK** (among others) investigated Clearview following complaints by privacy advocates. Each found Clearview in breach of the GDPR and issued enforcement actions. For

[127] Catherine Jasserand, 'Clearview AI: Illegally Collecting and Selling Our Faces in Total Impunity? (Part II)' (CiTiP Blog, KU Leuven, 5 May 2022) https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-ii/
[128] Ibid.
[129] van Haperen (n 67) 12.
[130] Jasserand (n 125)

example, France's CNIL in December 2021 declared Clearview's operations illegal and ordered the company to **cease processing French residents' data and delete existing data**.[131] When Clearview ignored this order, CNIL followed up with the maximum GDPR fine of €20 million in October 2022,[132] and later an additional €5.2 million penalty in 2023 for continued non-compliance.[133] Italy's Garante similarly imposed a €20 million fine in February 2022, and Greece's DPA levied another €20 million fine in July 2022.[134] The **Dutch DPA** (Autoriteit Persoonsgegevens), after its own investigation, announced in 2024 a fine of €30.5 million, emphasising that Clearview's creation of biometric identifiers from scraped photos had no legal justification and violated multiple GDPR provisions. These fines, among the highest available under GDPR, signal that **European regulators are deadly serious about enforcing data protection in the context of AI-driven biometric surveillance**.

It is worth noting a wrinkle in the UK: the Information Commissioner's Office (ICO) issued a £7.5 million fine and deletion order in May 2022, but Clearview appealed. In a controversial October 2023 decision, the UK's First-Tier Tribunal allowed the appeal, accepting Clearview's argument that its processing fell outside UK GDPR because the company was providing services solely to foreign (non-UK) law enforcement.[135] The ICO is **appealing** this ruling, arguing that Clearview itself, not the foreign police, was the controller processing UK individuals' data, thus UK law *should* apply.[136] This jurisdictional dispute illustrates challenges in applying GDPR to a company with no EU/UK establishment that *claims* to only serve overseas clients. Nonetheless, EU DPAs (unlike the UK tribunal) have maintained that GDPR does reach Clearview due to the "monitoring" of individuals in Europe (per GDPR Art . 3(2)): by collecting Europeans' images and creating profiles, Clearview is certainly monitoring behaviour in the eyes of EU law.[137]

---

[131] van Haperen (n 67) 12.
[132] Ibid 12.
[133] Ibid 12.
[134] Privacy International, *Challenge against Clearview AI in Europe* (Privacy International Legal Action, 27 May 2021) https://privacyinternational.org/legal-action/challenge-against-clearview-ai-europe
[135] Ibid.
[136] Ibid.
[137] Jasserand (n 125)

Compliance under the EU AI Act: The EU's Artificial Intelligence Act, which at the time of writing (2025) has been adopted (Regulation (EU) 2024/1689) but not yet fully in force, directly targets systems like Clearview's. Under the AI Act, **remote biometric identification systems** used in publicly accessible spaces for law enforcement purposes are classified as **unacceptable risk** and **prohibited** (Article 5(1)(d)–(f) of the AI Act), subject to narrow exceptions for serious crimes and only with judicial authorisation.[138] Clearview's tool enables law enforcement to perform biometric identification (albeit not always in real-time, but after-the-fact searching of images). If used for ongoing surveillance or broad identification in public, it would likely fall under this ban, reflecting a European consensus that such uses of AI inherently pose excessive risks to privacy and civil liberties. Even outside the outright ban, the AI Act treats **biometric identification systems in general as "high-risk" AI** when used by police or others (Annexe III, section 1).[139] This designation triggers a stringent compliance regime: providers of high-risk AI like Clearview's would have to undergo **conformity assessments** before putting the system on the EU market, ensure robust risk management and data governance, and implement oversight mechanisms.[140] For instance, high-risk AI providers must create detailed technical documentation and logging to facilitate audits, and design the system to allow human oversight and the ability to **override or shut down** the AI if necessary.[141]

Clearview AI, as it currently operates, would have extreme difficulty meeting the AI Act's standards. Its data is collected in violation of GDPR (hence not "lawful" or ethically sourced, which contravenes the AI Act's data governance requirements), and the system's **accuracy and bias** are questionable facial recognition algorithms are known to have higher error rates for women and minorities, which could lead to false matches and discriminatory outcomes.[142] Under the AI Act, if Clearview's system were evaluated, it might fail the requirement to eliminate or mitigate such bias to an acceptable level. Moreover, the **AI Act's transparency obligations** would require that

[138] Almada (n 83) 61-62.
[139] Ibid 63.
[140] Ibid 65.
[141] Ibid 66.
[142] Privacy International (*n* 132)

individuals be informed when they are subject to facial recognition in public (something fundamentally at odds with covert scraping and matching). In short, the AI Act would likely either **bar Clearview's current business model outright** or impose such heavy compliance burdens (e.g. prior authorisation by a legal authority for each use, mandatory bias testing, etc.) that Clearview could not operate in the EU. The Clearview case, in fact, was a catalyst in EU policy discussions it exemplified the dangers of unregulated facial recognition, prompting lawmakers to ensure the AI Act draws a hard line against pervasive biometric surveillance.[143]

Compliance under the Algorithmic Accountability Act (AAA): Unlike the EU, the United States lacks a comprehensive data protection law, and thus, Clearview's activities did not violate a GDPR-equivalent domestically. However, the proposed Algorithmic Accountability Act (introduced in Congress, with versions in 2019, 2022, and 2023) seeks to address some of these regulatory gaps. The **AAA would require companies to conduct impact assessments** for automated decision systems that are deemed "high-risk" or involve "critical decisions" about individuals' lives.[144] While primarily aimed at decisions in finance, employment, healthcare, housing, etc., it also covers systems that **profile or affect consumers in significant ways**. Clearview's facial recognition service, if used to make decisions (e.g. identifying someone as a criminal suspect), could fall under the scope of "automated decision systems" with critical impact on rights. The AAA would compel a company like Clearview to **evaluate and document** the algorithm's design and its impacts on privacy, accuracy, and potential bias.[145] For instance, Clearview would need to assess whether its facial recognition exhibits disproportionate error rates for certain demographics a well-documented issue with facial AI and whether its data practices (scraping billions of images) present privacy or security risks. The AAA also envisions **transparency requirements**, such as disclosing the use of such AI systems and publishing summary reports of these impact assessments.[146]

---

[143] Almada (n 83)
[144] *Algorithmic Accountability Act of 2023: Summary* (Senator Ron Wyden, May 2023) https://www.wyden.senate.gov/imo/media/doc/algorithmic_accountability_act_of_2023_summary.pdf
[145] Ibid.
[146] Ibid.

Under AAA, the **Federal Trade Commission (FTC)** would be empowered to enforce these rules.[147] Clearview could have been required to register its system and provide the FTC with the results of a bias audit and privacy impact report. If it failed to address identified risks (for example, if an audit found that the system misidentifies people of color at unacceptably high rates, or that the data collection was inherently privacy-invasive), the FTC could take action, potentially labeling the practice an "unfair or deceptive" act in commerce. In essence, AAA does not ban technologies outright but **mandates accountability**: companies must **justify that their AI systems are not unduly harmful or discriminatory**.[148] Had AAA been in force, Clearview's sweeping, secretive surveillance tool would have faced intense scrutiny – perhaps deterring its deployment in the first place or forcing substantial safeguards (like external auditing and limited scope). In reality, absent federal action, Clearview operated in a regulatory vacuum in the U.S., facing legal challenges only under narrower statutes like Illinois' Biometric Information Privacy Act (BIPA). The AAA thus represents the kind of oversight that could rein in cases like Clearview by **shifting the burden onto the AI provider to prove its system's lawfulness, fairness, and necessity**.

Ethical Implications: Fairness & Non-Discrimination: Clearview's facial recognition system raises stark fairness issues. Facial recognition algorithms have been shown to perform unevenly across different racial and ethnic groups, often less accurately identifying women and people with darker skin tones, due to training data biases and inherent algorithmic bias.[149] The ethical risk is **misidentification**: an innocent person (disproportionately from a minority group) might be flagged as a match and find themselves questioned or arrested because "the computer said so." Such false matches linked to biased algorithms have real victims, as seen in cases where flawed facial recognition led to wrongful arrests of Black men in the U.S. This implicates the ethical principle of non-discrimination and equality before the law. A tool that works *better* on some populations than others, and that could amplify existing policing biases (e.g. more surveillance of certain

[147] Ibid.
[148] Ibid.
[149] Privacy International (*n* 132)

communities), is ethically fraught. Moreover, fairness from a societal standpoint is violated by Clearview's approach of **treating everyone's face as fair game** for inclusion in a police database. Individuals who have never been suspected of any wrongdoing are nonetheless **implicated in a virtual lineup** whenever an officer runs a face search.[150] This erodes the presumption of innocence and can disproportionately affect communities that are over-policed or over-photographed.

- **Transparency:** Clearview's operations were covert by design; individuals generally had no idea their images were being collected and could be used to identify them. This lack of transparency is an ethical failing under principles of respect for persons and autonomy. People were denied the agency to choose whether or how their likeness is used. Transparency is also a precondition for accountability: if an AI system is secret, those subjected to it cannot contest or correct errors. Ethically, Clearview created an **imbalance of power** between the surveillers and the surveilled, without the latter's knowledge. In democratic societies, such opacity in a tool that can affect one's rights (being investigated or watched by police) is widely viewed as unethical. The GDPR's transparency requirements reflect this ethical stance, and Clearview's blatant flouting of those requirements (e.g. ignoring information duties and even deleting references to GDPR from its privacy policy) underscores a dismissive attitude toward the rights of individuals.[151]

- **Accountability:** The Clearview case highlights serious accountability gaps. When a law enforcement agency uses Clearview and makes a decision (to investigate or detain someone) based on the AI's output, who is accountable if something goes wrong? Is it the police officer, the department, or Clearview's developers? Ethically, any impactful algorithm should have clear lines of accountability. However, Clearview's setup encourages buck-passing: the company might claim it just provides leads, while police might over-rely on the AI ("the algorithm gave us a match"), a phenomenon known as **automation bias**.[152]

[150] Ibid.
[151] Jasserand (n 125)
[152] Almada (n 83)

Accountability is further muddled by Clearview's refusal to disclose its accuracy or allow external auditing. An ethical AI framework demands not just technical reliability but also that someone can be questioned and held answerable for errors. In Clearview's case, **regulators stepped in to enforce accountability,** essentially holding Clearview accountable under data protection law for the processing it initiated, but until then, the system operated with impunity, and individuals had virtually no recourse. This is ethically troubling: a powerful AI system affecting potentially millions was **operating outside the normal accountability structures** of checks and balances that bind government or even corporate activities.

- **Fundamental Rights and Human Dignity:** At a broader level, critics of Clearview argue that its system is **incompatible with fundamental rights** and human dignity. The idea that anyone can be identified anywhere in public, without any oversight or consent, has a chilling effect on rights like freedom of expression and assembly. Ethically, this veers into treating persons as mere objects of data faces to be tracked rather than individuals with agency and dignity. The European ethic of privacy holds that individuals should be able to go about their lives without constant identification. By compiling "**a database of everyone**," Clearview created the infrastructure for a pervasive surveillance state, which civil society groups find ethically abhorrent unless strictly controlled. This ethical stance was echoed by the **European Data Protection Board**, which in 2021 called for a ban on indiscriminate use of facial recognition in public spaces, stressing that the technology's threat to privacy and freedom outweighs its purported benefits. Clearview's case thus raises fundamental ethical questions about **the kind of society we want to live in** – one of ubiquitous surveillance or one that preserves anonymity in daily life. The overwhelming response, through an ethical and legal lens in Europe, has been that Clearview's approach is beyond the pale.

Role of Regulators: EU Data Protection Authorities (DPAs): As detailed, DPAs took coordinated enforcement action against Clearview. This was a notable example of regulatory

cooperation: privacy advocacy groups (like Privacy International, NOYB, etc.) filed complaints in multiple countries simultaneously,[153] and DPAs responded with investigations and sanctions. France's CNIL, Italy's Garante, Greece's Hellenic DPA, and the Austrian DPA all issued orders or fines by 2022–2023.[154] The French and Italian authorities hit the maximum fine allowed (€20 million) indicating the gravity of violations in their view. Greece imposed its highest fine ever (€20 million) for Clearview.[155] Austria's DPA, interestingly, found the processing illegal but initially issued no fine, likely due to procedural nuances or perhaps awaiting coordination (it left the door open to impose one later).[156] Germany, lacking a single federal DPA decision, had earlier seen the Hamburg DPA declare Clearview's practices unlawful in 2020 and order data deletion for a specific complainant.[157] Sweden's DPA fined a police authority for illegitimately using Clearview, showing that regulators were willing to enforce not just against Clearview itself but also against any *user* of the system that breached data protection law. Collectively, these regulators sent a clear message: **GDPR applies to facial recognition, and companies cannot sidestep European law by operating from abroad**.

- **Follow-through and Enforcement Challenges:** Regulators faced the practical challenge of enforcing penalties on a company with no EU presence. Clearview has, to date, **not paid the European fines** and largely ignored the deletion orders, prompting CNIL's additional penalties for non-compliance. This highlights a limitation in regulatory reach: DPAs can issue sanctions, but collecting on them or coercing compliance may require further legal tools (international agreements, seizure of assets if any come under EU jurisdiction, etc.). Nevertheless, regulators did not shy away from using their powers. They also leveraged publicity: each decision was made public, further shaming the company and warning off potential business partners. Another aspect of the regulatory role was the European Data Protection Board (EDPB), which facilitated a unified approach. While formal "joint

[153] Privacy International (*n* 132)
[154] Ibid.
[155] Ibid.
[156] Ibid.
[157] Ibid.

operations" under GDPR weren't needed since Clearview had no main establishment in the EU, the EDPB supported a task force for the exchange of information among DPAs. This case has arguably strengthened the hand of regulators by setting precedents on tricky issues (like extraterritorial application and biometric data interpretation).

- **FTC (USA) and State Regulators:** In the U.S., the Federal Trade Commission has authority to enforce against unfair or deceptive business practices. While the FTC did not take public action against Clearview (no fine or consent order so far), **advocacy groups filed a complaint urging the FTC to intervene** early in 2020, arguing Clearview's surreptitious data collection and false claims (e.g. that it had a First Amendment right to scrape data) constituted unfair business practices. The lack of immediate FTC enforcement could be due to legal uncertainty in the absence of a clear privacy law; labelling Clearview "unfair" would have broken new ground. However, pressure remained on the FTC, and in late 2021, the agency did issue warnings to a number of AI companies about biometric data practices (though without naming Clearview explicitly). It's possible the FTC chose to let the issue play out through state-level actions first. Notably, the **Illinois Attorney General and the ACLU** brought a lawsuit under Illinois' BIPA, which Clearview settled in May 2022. In that settlement, Clearview **agreed to stop providing its service to most private companies nationwide and to refrain from selling to any Illinois entity (including police) for a period of years**.[158] This is a kind of regulatory outcome via litigation: a state biometric privacy law was used to curb Clearview's operations. Additionally, other states like California and New Jersey opened investigations or banned law enforcement from using Clearview. Thus, a combination of state regulators, attorneys general, and civil suits created a patchwork response in the U.S., even as the FTC remained relatively quiet.

[158] American Civil Liberties Union, *ACLU v. Clearview AI* (ACLU, 11 May 2022) https://www.aclu.org/cases/aclu-v-clearview-ai#legal-documents

- **Role of Courts:** Besides regulatory agencies, courts have played a role. In the EU, the aforementioned UK Tribunal ruling (currently under appeal) is one judicial piece. In the U.S., aside from the Illinois case, there is ongoing multi-district litigation against Clearview for privacy violations. These cases test legal theories (e.g. does collecting faceprints from public websites violate individuals' rights under existing law?). So far, the clearest victories against Clearview have come from the regulatory front in Europe and the BIPA enforcement in Illinois, demonstrating that **specialised privacy laws with private rights of action (like BIPA) or strong regulatory regimes (GDPR)** are effective tools. The AAA, if enacted, would bolster regulators like the FTC to take more decisive action federally.

Practical Outcomes or Enforcement Actions: Clearview AI's encounter with regulators has resulted in substantial financial penalties (over €50 million in fines across Europe to date)[159] and legal orders to delete EU personal data from its systems.[160] Clearview's response, however, has largely been defiance. The company has not voluntarily complied with deletion or paid fines, and it remains to be seen whether EU authorities can practically force compliance (e.g. through international legal assistance or if Clearview ever has assets or personnel in Europe). One *practical* outcome is that **Clearview geofenced Europe in a technical sense** it purportedly stopped offering services to EU clients as the heat rose (and in fact, outside of law enforcement, it had already started limiting its client base). By 2022, Clearview claimed it was only working with U.S. and some foreign (non-EU) law enforcement and no longer serving private companies globally.[161] This retreat is partly attributable to enforcement pressure. Thus, while EU regulators couldn't directly shutter Clearview, they **succeeded in driving it out of their jurisdictions (at least formally)** and stigmatising its model.

- The UK tribunal ruling in Clearview's favour (if it stands) could complicate the practical impact in that jurisdiction, potentially allowing Clearview a loophole (serving foreign police

[159] Jasserand (n 125)
[160] Privacy International (*n* 132)
[161] ACLU (*n* 156)

but impacting UK residents' data). However, that ruling is isolated and on a point of law (scope of UK GDPR); it did *not* endorse Clearview's model as fair or compliant in substance. The ICO's persistence in appealing that case shows regulators' commitment to not leaving such loopholes unchallenged.[162]

- In the United States, Clearview's legal outcomes have constrained it somewhat. The Illinois settlement carries the force of a court order; violation of that could result in contempt of court or additional penalties. Furthermore, Clearview now faces ongoing oversight from a compliance monitor as part of the settlement. **Private sector use of Clearview in the U.S. has essentially ceased** due to that settlement and due to negative publicity, a major practical outcome since it prevents, for example, retail or banking industries from using the tool to identify people (something that was briefly explored by some entities). However, law enforcement use in the U.S. continues, albeit under increased scrutiny. On that front, some practical pushback came from police themselves: several large city police departments that had tried Clearview (often without higher-up approval) later banned it when legal concerns were raised. For instance, the NYPD and LAPD issued directives prohibiting officers from using Clearview or similar apps unofficially.

- One significant outcome is that Clearview has become a **cautionary tale in tech**. The reputational damage was severe. Clearview's brand is toxic in much of the world. This has dissuaded investors and acquirers, limiting the company's growth. In practical terms, even aside from direct regulation, Clearview's example *warned off* other would-be "data scrapers" or facial recognition-as-a-service startups from pursuing the same path without considering privacy implications. The case effectively demonstrated that "move fast and break things" with personal data can trigger legal backlash. For regulators, this serves as a success insofar as it is shaping industry behaviour through a deterrent effect.

[162] Privacy International (*n* 132)

- Finally, the enforcement saga around Clearview has spurred **public awareness and debate**. Media coverage of these enforcement actions and fines has educated the public about their rights (e.g., many European citizens learned they could request deletion of their images from Clearview via DPAs). This awareness is a soft outcome but important: informed citizens can better exercise their rights and support regulatory initiatives. It also pressures democratic governments to consider stricter rules on AI surveillance, which we see reflected in the advancement of the EU AI Act and various national proposals to ban facial recognition in certain contexts.

Lessons Learned & Implications for Future Regulation: *Data Protection Law's Strengths and Limits:* The Clearview case reaffirmed the power of laws like the GDPR to tackle novel AI harms. GDPR's broad principles of lawfulness, transparency, data minimisation, purpose limitation, etc., were more than sufficient to declare Clearview's practices unlawful.[163] This shows that existing privacy laws can adapt to AI challenges. However, the case also exposed **enforcement limitations**, especially with an overseas actor. The inability to physically enforce deletion or collect fines from a non-cooperative foreign company points to a need for **enhanced international cooperation**. One implication is to pursue international agreements or pressure through trade and diplomacy to enforce privacy norms globally. Another is considering innovative legal tools, for example, blocking orders at the network level (some have floated the idea of blocking Clearview's domains in the EU until compliance, though that raises its own issues). Regulators learned that a purely territorial approach is not enough in a global internet environment; hence, we may see more discussions on **reciprocity in data protection enforcement** (much like how competition law or anti-corruption law operates internationally).

- *Bolstering Deterrence:* Although fines were high, Clearview's case shows that for a venture-capital-backed startup with no EU assets, even €20 million fines can be shrugged off. Future regulatory strategy might include **personal liability** for executives in egregious

[163] Jasserand (n 125)

cases of unlawful AI use. For instance, GDPR allows for criminal referrals in some member states; if company officers faced personal sanctions, compliance might improve. Additionally, regulators might push for broad injunctions: under GDPR, they issued stop-processing orders, but enforcing those is tricky. Perhaps leveraging app stores or internet infrastructure (e.g. banning companies from government procurement or cloud services if they flout data laws) could be a path. Essentially, the lesson is that regulators must be creative and tough to actually *stop* a rogue AI operator, not just issue fines. This case will likely influence the EDPB's upcoming EDPB Enforcement Strategy, encouraging collective EU action for such scenarios.

- *AI Act and Precautionary Bans:* Clearview's case strongly influenced the EU AI Act's approach to biometric AI. The joint call by the EDPB and EDPS in 2021 for an outright ban on remote biometric identification was, in large part, a response to technologies like Clearview. The AI Act incorporated that to a degree (with law enforcement exceptions hotly debated). The lesson learned is that **some AI applications are so dangerous to rights that ex post facto enforcement (like GDPR fines) isn't enough upfront prohibitions or strict conditions are warranted**.[164] This represents a more precautionary regulatory philosophy. The implication for future regulatory strategy is to identify other such high-risk use cases early (for example, social scoring, AI-based lie detectors, etc.) and either ban or heavily regulate them before they become widespread. Clearview taught regulators that once the genie is out (billions of photos scraped and a product deployed), trying to put it back (through litigation) is extremely difficult. It's far better to **prevent** the harm by setting clear legal red lines in emerging AI regulation.

- *Algorithmic Accountability and Auditing:* Clearview also underscores the importance of independent **algorithmic auditing and accountability**. None of Clearview's claims about accuracy or lack of bias were subject to independent verification before deployment. Under

[164] Almada (n 83) 62.

an Algorithmic Accountability Act regime, this gap would be addressed by requiring impact assessments and possibly external audits.[165] The case strengthens the hand of those arguing for such laws: it's a vivid example where a company's unchecked deployment of AI had serious implications for society. Going forward, regulators are likely to insist on more **ex ante assessments** for AI systems, be it through law (as AAA proposes) or policy frameworks (the EU AI Act has a form of this via conformity assessments). Even absent formal law, we see a trend of public-private efforts to audit algorithms (e.g., NIST in the U.S. has a Facial Recognition Vendor Test that evaluates the accuracy and demographics of FR algorithms, something law enforcement could require vendors to participate in). Clearview's refusal to engage with such accountability mechanisms taught regulators and customers alike that **algorithmic claims should not be taken at face value;** they need verification.

- *Intersection of Privacy and Other Rights:* This case also highlights that governing AI is not just about data protection; it touches on surveillance law, law enforcement accountability, free expression, and anti-discrimination. A future regulatory strategy for AI needs to be **holistic**. For example, while GDPR handled the personal data aspect, some advocated for laws specifically restricting police use of facial recognition (some EU states and U.S. cities enacted such laws). We also saw human rights bodies weighing in. The lesson is that **multi-faceted governance** involving privacy regulators, law enforcement oversight bodies, legislators, and courts must converge to tackle AI issues. Relying on one area of law (like data protection) might leave gaps; a combination (data protection, sectoral bans and ethical guidelines) will be more effective. Clearview's saga likely will lead to stronger cooperation between different regulators (e.g., data protection authorities working with police oversight commissions or consumer protection agencies) when confronting AI that spans jurisdictions or sectors.

[165] Almada (n 83) 67.

- *Public Engagement and Ethical Norms:* Lastly, Clearview taught that public outcry and ethical norms can drive regulatory action. The case captured public imagination (for many, it was a first introduction to what AI can do with personal data). The **ethical consensus** that emerged that mass facial recognition is undesirable gave regulators the social license to come down hard. This implies that future regulatory strategy should involve **public dialogue about AI uses**. Regulators might increase efforts to educate and consult the public on AI deployments (the way some cities did with face recognition bans). When the public is informed and aligned on ethical limits, it empowers regulators to act boldly. In Clearview's case, regulators essentially enforced not just the letter of GDPR but its spirit, reflecting societal values about privacy and surveillance. For the future, one can envision regulatory bodies explicitly incorporating **ethical impact assessments** in their decisions (some DPAs already refer to fairness and ethics in their analyses). The lesson is that law and ethics in AI governance are intertwined, and effective regulation will continue to require both legal rigour and moral clarity.

In conclusion, the Clearview AI case demonstrates both the challenges and the indispensability of robust legal frameworks like the GDPR (and upcoming AI Act) in curbing the excesses of AI-driven decision-making. It also highlights why proposals such as the Algorithmic Accountability Act are gaining traction – to extend such governance to jurisdictions (like the U.S.) where gaps remain. Clearview's story will likely be cited for years as a test case of AI regulation: a scenario where **technological capabilities sprinted ahead, but legal and ethical oversight eventually caught up**, albeit needing to further strengthen its reach to fully remedy the harm.

6.2 SCHUFA: Automated Credit Scoring under Article 22 GDPR

Case Overview: SCHUFA Holding AG (commonly just "SCHUFA") is Germany's largest credit bureau, which computes credit scores for individuals. These scores – generated through proprietary

algorithms using personal data such as an individual's payment history, outstanding loans, court records of defaults, and sometimes demographic information are used by banks, landlords, and other businesses to make decisions about creditworthiness. In practice, a **SCHUFA score can determine** whether someone is granted a loan, gets a mortgage, can lease a car, or even sign a mobile phone contract. For years, SCHUFA's operations typified an automated decision-making system: an **algorithm produces a numeric score** that has a significant impact on people's financial opportunities. The legal question that arose (and which culminated in a landmark court ruling in 2023) was whether this practice falls under GDPR's Article 22 on automated decision-making, and if so, whether it is being conducted lawfully.

Legal Compliance under GDPR (Article 22 and beyond): Article 22(1) GDPR provides that individuals have **"the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."** In simpler terms, important decisions about people shouldn't be made solely by algorithms, unless an exception applies. For a long time, there was debate about whether a credit score itself is a "decision." SCHUFA maintained that it only provides a score to a lender, and the *lender* makes the actual decision to grant or deny credit (often with human involvement). On that reasoning, SCHUFA viewed its scoring as a preparatory activity, not the final automated decision. This interpretation was upended by the **Court of Justice of the EU (CJEU)** in *OQ v. SCHUFA Holding AG* (Case C-634/21), judgment of 7 December 2023.[166] The CJEU ruled that **SCHUFA's act of generating a credit score by automated means *does* qualify as automated decision-making** within the meaning of Article 22, *because* that score is used with significant weight by third parties in deciding on contracts.[167] In other words, even if SCHUFA itself does not say "loan approved or denied," the score "guides decisions about various aspects of an individual's life" and thus materially affects them.[168] The Court took a broad view of what counts

[166] Center for Information Policy Leadership, *Decoding Responsibility in the Era of Automated Decisions* (Hunton Andrews Kurth, Oct 2024)
[167] Ibid.
[168] Almada (n 83) 55.

as a "decision" it includes not only the final act (e.g. a loan rejection letter) but also intermediate outputs like a score that have a **decisive influence** on that outcome.[169]

Crucially, the CJEU clarified that Article 22(1) is a **prohibition in principle** on such automated decisions; it's not merely a right that individuals must invoke.[170] This means companies cannot engage in solely automated significant decisions unless they fit one of the narrow exceptions in Article 22(2). The three exceptions are: **(a)** necessary for entering into or performing a contract with the data subject; **(b)** authorized by law (with safeguards); or **(c)** based on the data subject's explicit consent.[171] In SCHUFA's context, typically no explicit consent is obtained from individuals for scoring (consent in such scenarios would likely not be freely given, as credit is conditional on it). There isn't a specific German law that authorizes private credit scoring and provides safeguards (though sectoral regulations exist, they don't explicitly override GDPR's default rule). That leaves **"necessary for a contract" (Art.22(2)(a))** as the plausible basis. Lenders might argue that automated scoring is necessary to decide whether to enter a credit contract with the individual. The CJEU did not categorically confirm that credit scoring meets this necessity test, but implied that if it were to be used, that exception would be the one to examine.[172] However, even if an exception applies, Article 22(3) mandates that certain **safeguards** be in place: at minimum, the right to obtain human intervention, to express one's viewpoint, and to contest the decision.[173] The Court underscored that these safeguards must accompany the process from the start, not just as a theoretical option.[174]

For SCHUFA and its business partners, compliance now requires adjustments. **If relying on the "contract necessity" exception**, credit scoring must be coupled with measures like giving individuals an opportunity to request a human review of an automated unfavourable outcome.[175] In

[169] Ibid 56.
[170] Ibid 56.
[171] Ibid 57.
[172] CIPL (n 164)
[173] Almada (n 83) 57.
[174] Ibid 57.
[175] Ibid 58 .

practice, this might mean, for example, if someone is denied a loan due to a low SCHUFA score, they should be informed of that fact and allowed to ask the lender for a manual reconsideration (which might involve looking beyond the score at their circumstances). It could also mean SCHUFA has to structure its service to enable those rights, e.g. providing channels for queries or objections from data subjects about their score. The **CJEU's ruling** shifts responsibility: it "falls on the credit reference agency rather than just on the lender" to comply with Article 22.[176] In other words, SCHUFA cannot hide behind lenders; it is a controller making automated decisions and must ensure the processing is legal.

Beyond Article 22, GDPR's **transparency and access provisions** come into play. Article 15(1)(h) GDPR specifically gives data subjects the right to request "meaningful information about the logic involved" in automated decisions. In the case that sparked the CJEU ruling, the individual (OQ) had indeed made an **access request to SCHUFA asking for an explanation of the score**.[177] SCHUFA refused, citing trade secrets in its algorithm. The CJEU case did not fully resolve how to balance trade secrets vs. transparency (that issue was partly addressed in a parallel case on data retention by credit agencies). However, the clear implication is that **SCHUFA will need to provide more information** than it historically did. A generic statement like "your score is X because of our proprietary algorithm using your credit history" may not suffice. Individuals may not be entitled to the exact algorithm or formula (as per Recital 63 and trade secret protections), but **at least the main factors that negatively influenced their score** should be disclosed. For compliance, SCHUFA and similar agencies might develop templates: e.g., informing someone, "The factors that contributed to your score include a history of missed payments and lack of recent positive credit history, weighted in such-and-such way" enough to be meaningful without revealing the full formula. German law (§34 of the Federal Data Protection Act) already had provisions for data subjects to obtain their score and some information annually, but the CJEU ruling elevates the requirement to ensure an explanation of the *logic*, not just the output, is accessible.

[176] CIPL (n 164)
[177] Ibid.

In summary, under GDPR as now interpreted, SCHUFA's automated scoring *per se* is not forbidden, but it **must operate within a tighter legal framework**: either stop purely automated scoring or implement one of the Art. 22(2) exceptions with robust safeguards and transparency. Failure to do so could mean every issuance of a score is an unlawful automated decision, opening SCHUFA (and possibly its partner lenders as joint controllers) to GDPR complaints and fines.

Compliance under the EU AI Act: Credit scoring systems are explicitly identified as a high-risk AI use case in the AI Act. Annexe III of the Act (as adopted) lists among high-risk AI systems those used in **"access to essential private services,"** which includes creditworthiness evaluations for loans.[178] This means that an AI system that assesses a person's credit risk, like the algorithm SCHUFA uses, will be subject to **ex ante compliance requirements** once the AI Act applies. SCHUFA could be considered either a "provider" of a high-risk AI system (if we treat the scoring algorithm as a product it offers to lenders) or at least a "user" deploying such a system in its own decision-making process. In either case, major obligations follow:

- **Data Quality and Bias Mitigation:** The AI Act will require that training data for high-risk AI is relevant, representative, and free of errors *to the best extent possible* (Article 10). For SCHUFA, this could mean ensuring the datasets used to develop or calibrate the scoring model are statistically sound and do not systematically disadvantage protected groups. If, for instance, the model uses postcode as a factor (which could correlate with socio-economic status or ethnicity), SCHUFA might need to justify this or show it doesn't result in unfair bias. The Act's emphasis on avoiding discrimination means credit scoring AI should be vetted for disparate impact. Thus, under AI Act compliance, one might expect **algorithmic impact assessments** similar to those envisaged by the AAA: testing whether the score predictions have any bias against, say, residents of certain neighborhoods or against young people, etc., and adjusting the model if they do.

[178] Almada (n 83) 59.

- **Transparency & Information to Users:** The AI Act will require that users of the AI system (in this case, lenders using SCHUFA's system, or SCHUFA's own staff interpreting scores) are provided with information about the system's capabilities, limitations, and appropriate use (Article 13). SCHUFA will likely have to furnish documentation about what the score means, its accuracy, and how it should (and should not) be used in decision-making. Moreover, if a consumer is subject to an AI-driven scoring, the AI Act's provisions (in conjunction with GDPR) might strengthen their right to be informed that an AI was used in that process.

- **Human Oversight:** The AI Act's Article 14 mandates that high-risk AI systems be designed with human oversight in mind. Practically, SCHUFA's system might need to include features or tools that allow a human reviewer (either at SCHUFA or the lender's side) to understand how a particular score was derived, or to override aspects of it if necessary. This dovetails with GDPR's requirement of human intervention on request.[179] The AI Act could push for a more proactive approach: not just reacting to individual requests, but **building systems such that humans are continuously in the loop or at least monitoring the AI's outputs for anomalies**. For example, if the AI suddenly drops someone's score dramatically due to an outlier data point, a human oversight mechanism might catch that and flag it.

- **Conformity Assessment and CE Marking:** As a high-risk AI provider, SCHUFA (or the vendor of its scoring software) would have to undergo a conformity assessment before the system can be put into use (Article 16 and following). They would effectively need to **audit the system's compliance** with all the AI Act requirements (possibly involving external notified bodies if required). Once compliant, the system gets a CE marking indicating it meets EU AI standards. This process compels a thorough check on aspects like cybersecurity (to ensure the score can't be manipulated), performance (the system should be as accurate as claimed), and robustness (e.g., not overly sensitive to minor data changes).

[179] Ibid.

The AI Act thus adds a **layer of procedural and technical compliance** beyond GDPR's individual rights focus. For SCHUFA, aligning with the AI Act might require changes such as producing a *detailed technical dossier* for regulators, instituting continuous monitoring of the model's performance, and possibly registering the system in an EU database for high-risk AI. In effect, the **adequacy of governance is enhanced**: GDPR handles personal data and fairness to the individual, while the AI Act addresses the algorithm's quality and the provider's diligence. One important interplay: The CJEU's broad take on Article 22 means SCHUFA must allow human intervention and contestation; the AI Act complements this by requiring the system be designed so that human intervention is actually effective (e.g., no "black box" models that humans can't interpret at all).[180]

Compliance under the Algorithmic Accountability Act (AAA): In the U.S., credit scoring has long been regulated by the Fair Credit Reporting Act (FCRA), which gives consumers rights to see their credit reports and correct errors, and by the Equal Credit Opportunity Act (ECOA), which prohibits credit decision discrimination on protected grounds. However, neither law specifically forces an examination of the *algorithms* or *models* used for credit scoring. The proposed Algorithmic Accountability Act would step into this gap. Under AAA, credit scoring systems would likely be classified as automated decision systems making "critical decisions" about consumers' finances.[181] As such, the law would require any company like SCHUFA (if it were under U.S. jurisdiction) or any lender developing its own scoring model, to perform an **Automated Decision System Impact Assessment**.

This assessment would entail: evaluating the scoring algorithm for **accuracy, fairness, bias, discrimination, privacy, and security**.[182] For instance, the company would need to analyze whether the model produces systematically lower scores for certain minority groups even when controlling for legitimate credit factors and if so, address that (or justify it in terms of business necessity, though a biased outcome could run afoul of ECOA regardless). They'd also assess

[180] Ibid 60.
[181] Algorithmic Accountability Act of 2023: Summary (n 142).
[182] Ibid.

privacy impacts e.g., is the data collected for scoring proportionate and secured? The AAA would require companies to **document these findings and report certain information to the FTC**.[183] It also pushes for a level of **transparency** consumers should be able to know when an algorithm was used in a decision about them and have some ability to get an explanation or appeal. In fact, the AAA's ethos resonates with Article 22 GDPR, but in a more proactive auditing manner rather than an individual rights manner.

If AAA were in place, an outfit like SCHUFA would be legally compelled to **proactively search for biases** in its scoring model and take corrective action (e.g., modify the algorithm or include additional safeguards) before problems manifest widely. It would also have to share the impact assessment (or a summary) with the FTC and possibly the public. In effect, AAA would make algorithmic accountability a **routine compliance matter** for credit scoring, rather than relying on individuals to challenge decisions after the fact. Since AAA empowers the FTC to promulgate detailed rules, it could even require specific metrics or thresholds (for example, requiring that the error rates or denial rates between different demographic groups not exceed a certain ratio unless justified). This could bring credit scoring practices under a degree of uniform federal oversight that currently doesn't exist (right now, oversight is more reactive, through litigation or enforcement if discrimination is suspected).

One challenge would be how trade secret protections interact with AAA's transparency goals. Credit scoring companies guard their formulas closely (just as SCHUFA did). The AAA likely would not force publication of the exact algorithm, but it might require disclosure of features used and steps taken to ensure fairness. This is analogous to how, in lending, banks must provide adverse action notices with reasons for denial. AAA could mandate a more technical disclosure. The **role of regulators like the FTC and CFPB** would be crucial: they would have to review these assessments and could investigate if, say, a company consistently shows disparities in their

[183] Ibid.

outcomes. The CFPB, in fact, has shown interest in algorithmic scoring, warning companies that "black box" models are not a defence against liability for disparate impact.

In sum, AAA would complement existing U.S. financial laws by injecting a systematic algorithm audit requirement, which is more in line with modern AI governance thinking. For future regulatory strategy, combining AAA's approach (preventive auditing and FTC oversight) with GDPR-like individual rights would create a robust framework. The SCHUFA case underscores why: had such measures been in place earlier, perhaps consumers wouldn't need to fight in court for basic transparency, the onus would have been on the scorer to ensure fairness from the start.

Ethical Implications: Fairness and Non-Discrimination: Automated credit scoring sits at the intersection of ethics and economics. Ethically, fairness demands that individuals are judged on relevant, accurate factors and that the process does not *unjustly discriminate*. A key concern is **proxy discrimination**: credit algorithms might use inputs that correlate with protected characteristics (like race or gender) without explicitly including them. For example, neighbourhood data, educational background, or even first names could introduce bias. If a scoring model yields consistently lower scores for certain minority communities or other protected groups, that's an ethical red flag even if the data points used seem facially neutral. There is also the issue of **socio-economic fairness**: credit scores can entrench disadvantage. Ethically, one might argue for giving individuals a chance to **explain special circumstances** (e.g., a past medical debt that caused a temporary financial issue) rather than purely scoring them by an unforgiving formula. Relying solely on an algorithm might ignore context and human nuance, which is why GDPR's safeguard of human intervention aligns with the ethical principle of fairness, allowing a holistic review. The SCHUFA case highlighted that point: treating a person as a number without recourse was deemed inappropriate. Going forward, ethically "fair" AI in credit should incorporate measures to avoid discrimination (like algorithmic fairness techniques) and include provisions for **individual assessment** when needed (no one should be irrevocably condemned by a past data point if circumstances have changed).

- **Transparency and Explainability:** Ethically, individuals have a right to know **why** a decision affecting them was made. With credit scoring, the algorithms are often **opaque** trade secrecy is invoked to keep the scoring formula secret, but from an ethical standpoint, this opacity is problematic. It leaves individuals in the dark about how to improve their situation or challenge a result. The notion of **respect for persons** entails giving people insight into decisions that profoundly affect them. In the SCHUFA context, many people only see a score and a vague explanation ("score calculated based on your credit history"), which might not be meaningful enough to contest or learn from. Ethically, this is insufficient. The CJEU's insistence on Article 22 rights and meaningful information echoes an ethical stance: **algorithmic decisions should not be completely inscrutable** to those who are subject to them. There's also a **social transparency** aspect; credit scores affect societal opportunities, so some argue the logic behind them should be transparent to public scrutiny to ensure it's fair and rooted in rational, non-exploitative criteria. Ethical AI frameworks often list "transparency" or "explainability" as key principles; in practice, for credit scoring, this might mean using more interpretable models or at least providing tools to explain individual outcomes (e.g., listing key factors that impacted a particular score).

- **Autonomy and Consent:** Another ethical issue is that individuals rarely have a **choice** in being scored. If you want a loan, you must accept that a credit score will be consulted. This raises questions of consent and autonomy. Under GDPR, consent is not really viable here, which underscores that some processing is imposed on individuals by economic necessity. Ethically, if something is non-consensual, it ought to have a strong justification and safeguards. The use of one's personal financial data and history to algorithmically score them is justified by the need for lenders to manage risk, arguably a legitimate aim. But the ethical quid pro quo is that individuals deserve robust rights in return (accuracy, recourse, etc.). If a person cannot opt out of the system (unless they forgo credit entirely, which in modern society is nearly impossible), then the system owes them fairness and explanation.

This is related to the concept of **power imbalance**. Credit bureaus hold significant power over consumers, so ethical practice calls for checks on that power (like regulation enforcing transparency and accuracy).

- **Accountability and Redress:** Ethically, when an algorithm like SCHUFA's makes a mistake or produces an unjust outcome, it must be possible to correct it and hold someone accountable. In the past, consumers often found it difficult to get errors in credit reports fixed or to challenge a low score because the process was opaque and cumbersome. That's an ethical failing that systems that affect lives should have clear accountability channels. With automated scoring, one worry is that humans in institutions **over-rely on the algorithm** ("computer says no") and abdicate their responsibility to make their own judgments or to question the algorithm. Ethically, this is problematic because it reduces accountability the blame gets pinned on the algorithm ("it's just how the system works"). GDPR's requirement for human intervention partly aims to restore accountability by ensuring a human is responsible at some stage. Also, we must consider **accountability for the design** of the algorithm: if it systematically undervalues a certain group, the designers and deployers should be accountable for that discrimination. In summary, ethical governance of such AI requires clear lines of responsibility and effective means for individuals to seek redress (like correcting wrong data or appealing a decision). The SCHUFA litigation itself is a form of seeking accountability the individual went to court to force the system to explain itself. That signals that normal accountability mechanisms were insufficient, hence the ethical need to improve them (through law, oversight bodies, etc.).

- **Implications for Social Justice:** A broader ethical lens sees credit scoring as part of a system that can perpetuate social inequality. Those who are already disadvantaged (low-income, less financial literacy, victims of past discrimination) may have lower scores and thus face higher barriers to economic participation (higher interest rates, difficulty renting apartments, etc.). Automated scoring can amplify these issues if not carefully

managed. There's an ethical argument for building **fairness constraints** into the system for example, adjusting models to be more forgiving of certain one-time hardships, or including positive data (like rent payment history) to give a fuller picture of creditworthiness for those who lack traditional credit. Some ethicists propose concepts like a "right to explanation" and a "right to better credit" through constructive feedback (e.g., telling people how they could improve their score). While not legal requirements, these ideas come from an ethical desire to make the system *less punitive and more rehabilitative*. In that sense, transparency is not just about explaining a past decision, but enabling future improvement which is an ethical good (empowering individuals to take control of their financial reputation).

In essence, the ethical implications highlight that **credit scoring AI must be fair, transparent, and accountable to those it scores**. The SCHUFA case has forced these ethical considerations to be translated into legal mandates. Ethics and law converge here: what was ethically troubling (secret, unchallengeable algorithms making important decisions) is now recognised as legally unacceptable in Europe.

Role of Regulators: Data Protection Authorities (DPAs): The SCHUFA case that reached the CJEU began with a complaint to a DPA, specifically, the data subject OQ complained to the **Hamburg Data Protection Authority** about SCHUFA's refusal to provide full information and about the practice of scoring.[184] The Hamburg DPA initially rejected the complaint (aligning with SCHUFA's view that Article 22 didn't apply). This illustrates that regulators themselves were divided or cautious on interpreting the law. However, once the matter was referred to the CJEU and a ruling obtained, DPAs are now bound to follow that interpretation. The CJEU's decision effectively gives DPAs across Europe a green light (and a directive) to treat credit scoring as within their purview under Article 22. We can expect DPAs to become more active in this area. For example, DPAs might issue **guidance or a collective EDPB statement** on automated credit scoring

[184] CIPL (n 164)

in light of the judgment, clarifying expectations (such as requiring that individuals be informed of their right to human review).

Going forward, if SCHUFA or other credit agencies do not adjust their processes, they could face enforcement. A likely scenario is DPAs conducting an audit or requesting information from SCHUFA on how it will comply. Since SCHUFA is based in Germany, the **lead DPA** would be the Hessian Commissioner for Data Protection (as SCHUFA's headquarters are in Wiesbaden, Hesse). The Hessian DPA or the German Datenschutzkonferenz (the body of all German DPAs) might engage with SCHUFA to update the industry's code of conduct. Notably, there *was* a code of conduct for credit agencies in Germany (under Art. 40 GDPR, approved by DPAs) that, for instance, set rules on how long data like insolvencies can be kept. The CJEU's joined Cases C-26/22 and C-64/22 (decided the same day as SCHUFA case) dealt with aspects of that code (retention of erased public debt records).[185] The Court held, inter alia, that data subjects must have the right to full judicial review of DPA decisions on such matters.[186] This empowers individuals to challenge DPAs if they feel the regulator isn't strict enough. So DPAs, knowing their decisions can be fully appealed, will likely be diligent in applying Article 22's new interpretation.

DPAs also have a role in the **interplay with consumer protection**. Some countries might see consumer protection agencies working with DPAs (since credit scoring affects consumer rights). A regulator like the **European Data Protection Supervisor (EDPS)** might weigh in for EU institutions or in a convening role though credit is mainly national, EDPS has an interest in consistent application across Europe and might include credit scoring in broader AI oversight discussions.

- **Financial and Consumer Regulators:** Apart from DPAs, credit scoring is overseen by financial regulators to some extent. For example, in Germany, the Federal Financial Supervisory Authority (BaFin) doesn't directly regulate SCHUFA (as it's not a bank), but it

[185] Ibid.
[186] Ibid.

regulates banks that use SCHUFA. BaFin might integrate the GDPR/Schufa ruling into its guidance to banks, e.g., advising banks that they should not rely solely on SCHUFA scores without human review, or that they must ensure any automated risk assessments they use comply with data protection laws. In the EU more broadly, institutions like the European Banking Authority (EBA) could update their guidelines on loan origination to incorporate data protection considerations (the EBA has spoken about the use of AI in creditworthiness and the need for fairness). There could be a collaborative approach: DPAs ensuring data rights, financial regulators ensuring systemic fairness and stability.

- **Courts and Judicial Oversight:** Regulators in this context include courts when adjudicating disputes. The German administrative courts were involved in the SCHUFA case and will be again when it returns from the CJEU to the national level. If SCHUFA or lenders push back against DPA enforcement (for instance, if issued an order or fine), courts will be the final arbitrator. Thus, ongoing judicial oversight is part of the regulatory ecosystem. The CJEU's strong stance suggests courts will uphold strict compliance. One practical impact: companies may approach DPAs to seek clarity (or even pre-approval of certain practices) to avoid litigation. DPAs sometimes use instruments like "regulatory sandboxes" or consultations for AI. We might see something like that if an industry player proposes a novel way to implement automated scoring with compliance; they might seek the DPA's informal nod that it meets the requirements.

- **In the United States:** Though the GDPR/AI Act doesn't apply in the U.S., it's informative to consider how regulators there deal with similar issues. The **Consumer Financial Protection Bureau (CFPB)** and the **Federal Trade Commission (FTC)** are key regulators. The CFPB enforces the FCRA and ECOA. In recent years, the CFPB has shown interest in algorithmic credit scoring (especially alternative credit scoring using big data). They have warned companies using AI that they will be held accountable for algorithmic bias under ECOA's disparate impact doctrine. The FTC can act if a credit scoring method is found to be

unfair or if a company misrepresents what its score means (deception). If an AAA were passed, the FTC would have direct regulatory tasks as described. State regulators also play roles, e.g., New York's Department of Financial Services has looked at credit scoring in insurance (which is analogous, as insurers use credit scores for underwriting, raising bias concerns).

There's also an increasing role of **civil rights organisations** as quasi-regulators: for instance, the National Fair Housing Alliance and other advocacy groups have sued or pressured companies for algorithmic bias (including credit algorithms). While not "regulators" in the formal sense, their actions influence regulatory agendas and outcomes (the way EPIC and others did with HireVue in the next case, or how ACLU did with Clearview). For example, if the NAACP or a similar organisation were to publish research that SCHUFA-like systems in the U.S. disadvantage Black consumers, that could spur regulatory or legislative action.

- **AI Act Oversight Bodies:** When the AI Act comes into force, each Member State will designate or establish a **Market Surveillance Authority** for AI. These bodies will oversee compliance with high-risk AI systems. In the context of credit scoring, that could mean checking that providers of such AI (like a fintech offering credit score algorithms to banks) meet the obligations. If SCHUFA's scoring system undergoes changes or a new system is introduced, this authority could audit the system's documentation, demand proof of risk mitigation, etc. The AI Act also empowers these authorities to order the recall or cessation of AI systems that pose serious risks. In extreme cases, if a credit scoring AI was found to be irredeemably biased or unsafe, an authority could halt its use (this would be an extraordinary step, likely only if it flagrantly violated requirements). The synergy between DPAs and AI Act authorities will be interesting; ideally, they will collaborate, since one focuses on data rights and the other on technical robustness and compliance. Some countries may merge these functions or have one agency handle both (e.g., France's CNIL might also take on AI Act duties).

Practical Outcomes or Enforcement Actions: The immediate outcome of the CJEU's SCHUFA ruling is that **SCHUFA (and by extension, similar agencies) must adjust their practices**. Although SCHUFA publicly stated that its scoring was compliant and necessary, the legal reality is that it must now either ensure that an exception under Article 22(2) applies with safeguards or stop pure automation. We have yet to see SCHUFA's official response (as of early 2025), but possible changes include: providing more disclosures in credit reports, implementing an internal mechanism for manual review upon request, and perhaps working with lenders to guarantee they handle the "human intervention" part when using scores. For instance, SCHUFA might include a notice with every score it returns: "This score was generated automatically. If you, as the lender, plan to reject the application based on this score, be advised that the applicant has the right to seek human review of the decision." This nudges lenders to comply as well.

- Lenders themselves may adapt processes. Some banks might now, as a policy, **avoid sole reliance on SCHUFA scores**. They could mandate loan officers to always double-check critical cases, or to consider supplementary information (like a conversation with the applicant or additional documents) before final rejection. In effect, credit decisions in Germany might become a bit less automated, injecting a measure of human discretion. This could slow down some decisions or require retraining staff, but it could also improve the perceived fairness of the process.

- For individuals, one tangible outcome is **improved access to information**. Post-ruling, SCHUFA can expect an uptick in data access requests and challenges from consumers armed with the knowledge that they have these rights. Consumers might request not just their raw score, but an explanation of it. If SCHUFA or lenders fail to provide it, we might see further complaints and even lawsuits. Consumer organisations in Germany, such as the Stiftung Warentest or consumer protection centres, might actively assist people in making use of their new rights (perhaps publishing templates for requests or running awareness

campaigns). Over time, this could lead to a more transparent environment where people routinely get feedback on their credit applications beyond "you were denied, sorry."

- Another outcome is potential **improvement in data accuracy**. Since individuals can contest decisions and scores, SCHUFA will be under pressure to ensure its data (fed from banks, courts, etc.) is correct. Any erroneous entry that drags down a score can lead to complaints and possibly compensation claims. So SCHUFA is likely to invest in better data quality controls. They may also reduce the weight of factors that are controversial or less predictive to avoid challenges e.g., if including some factor leads to many complaints or appears unfair, they might drop it to maintain trust and reduce friction.

- At a systemic level, the case might spur the **modernisation of credit scoring models**. With AI Act requirements coming, credit bureaus could explore using more sophisticated AI that, for example, provides reason codes for decisions (some AI models can pinpoint which inputs contributed most to an output for a given individual). If they can present those reasons to consumers (e.g., "Missed payments on account X and lack of recent credit history lowered your score"), it satisfies both GDPR and customer relations. Traditional credit scoring (like linear or logistic regression models) already provides reason codes, but more complex machine learning models could too if designed with explainability. We could see innovation in **explainable AI for credit scoring** as a response to these regulations.

- Enforcement actions to watch: if SCHUFA or others do not comply, DPAs might issue fines or orders. Given the prominence of SCHUFA, any enforcement would be high-profile. A failure to implement Art. 22 safeguards could theoretically lead to fines up to 4% of SCHUFA's turnover (as it'd be a grave GDPR violation). But more likely is a negotiated solution or warning before it reaches that stage. On the positive side, if SCHUFA adapts well, it could become a best practice example, and regulators might publicise that.

Conversely, any dragging of feet could lead to swift regulatory action now that the law is clarified.

- The ruling also has a **broader European impact**. Other EU credit bureaus (TransUnion's subsidiary in some countries, CRIF in Italy, Experian in the UK, though post-Brexit not under GDPR, etc.) are surely reviewing their processes. They may proactively implement similar changes EU-wide. Even outside the EU, the case is noted: for instance, Brazil's data protection authority (ANPD) and India's upcoming data protection law have comparable clauses on automated decisions, which they might interpret similarly, influenced by the CJEU's reasoning. So the case contributes to a global precedent that automated credit scoring should incorporate human oversight and transparency.

- Interestingly, the case might indirectly affect **fintech innovations**. Some fintech lenders use fully automated AI for credit decisions (sometimes using non-traditional data like phone bill payments or even social media data in some places). They will need to consider how to reconcile that with Article 22 in Europe. It might slow the adoption of "black box" credit scoring in the EU (which is likely a good thing for fairness, but companies may grumble about lost efficiency). Fintechs might need to build hybrid models (AI + human final check) or ensure explicit consent (though consent is fraught in credit contexts, as mentioned). Some may lobby for national laws under Art. 22(2)(b) to explicitly allow certain automated credit evaluations with safeguards, though passing such laws could be controversial (it would seem like weakening privacy rights, which is a hard sell politically).

In short, practically, the SCHUFA case forces a recalibration: making sure the convenience of automation does not override individuals' rights. If implemented well, the outcomes could be positive – fairer decisions, more engaged consumers, and credit algorithms that are more transparent and perhaps more accurate by virtue of being scrutinised. There may be some cost

(more manual work for lenders, possibly slightly slower lending decisions), but regulators evidently believe that is a worthwhile trade-off for protecting individuals.

Lessons Learned & Implications for Future Regulation: *Clarifying the Scope of "Automated Decisions":* One lesson from the SCHUFA saga is that legal ambiguity in definitions can allow practices to develop that skirt the intention of the law. Pre-2023, some thought an automated *intermediate* assessment (like a score) might not trigger Article 22 since a human eventually rubber-stamps it. The CJEU's clarification eliminated this grey zone: if it significantly affects people, it counts.[187] Regulators and lawmakers in future should ensure definitions in law capture the realities of AI usage. For instance, as AI systems get more complex (like multi-step decision pipelines), we need clear criteria for what constitutes a "decision" and "solely automated." The ruling sets a principle that will guide that: focus on **effect and substance, over form,** an ostensibly "advisory" algorithm can be de facto decisive. This lesson will likely inform interpretative guidance for other contexts (e.g., AI hiring tools could be treated similarly if an algorithm's score heavily influences the human hiring decision).

- *Rights are Only Effective if Enforced:* The case demonstrated that having a right "not to be subject to automated decisions" meant little until someone enforced it. Many companies likely assumed Article 22 wouldn't be actively applied or would be interpreted narrowly. Now that it has been enforced at the highest level, it sends a message that these GDPR rights are real and must be baked into AI system design. Future regulatory strategies should include **awareness campaigns** so that data subjects know and exercise their rights, and **regular audits** by DPAs to ensure companies aren't quietly making solely automated decisions without safeguards. Another implication is that regulators might push for **organisational measures,** for example, requiring companies to train staff about handling requests for human review and to document when such reviews happen (so they can prove compliance). Essentially, making Article 22 operational in organisations is the next

[187] Almada (n 83) 56.

challenge. The EDPB might update its guidelines on automated decision-making (which existed in draft) to reflect the CJEU decision, giving practical advice.

- *Human Oversight Must Be Meaningful:* The lesson "add a human in the loop" can be done badly or well. A token human who always trusts the algorithm adds no value. Regulators will likely look at how to ensure the human intervention is **substantive**. This could become part of the risk-based approach in AI regulation: e.g., the AI Act's requirement that humans can "understand and properly oversee" the AI.[188] It's not enough to have a checkbox saying "human review available"; the quality of that review matters. In credit, that might mean training loan officers to override scores in appropriate cases and monitoring how often they do so (if they never do, maybe the human oversight isn't genuine). So future regulatory strategy might involve **auditing the human aspect**: DPAs could ask, "In the past year, how many automated denials were overturned upon request for human intervention?" If the answer is zero, that might be a red flag. The broader implication is that as AI pervades decisions, regulators may need new techniques to assess compliance, not just reviewing code or privacy policies, but reviewing organisational behaviour and decision logs.

- *Interdisciplinary Cooperation:* We touched on how different regulators will coordinate. The lesson is that AI in important fields (credit, employment, etc.) cannot be siloed. Data protection authorities, consumer protection agencies, anti-discrimination bodies, and sectoral regulators should share expertise and align their actions. For instance, a DPA could handle the transparency and consent side, while an equality body might address whether the outcome pattern is discriminatory. In an ideal future regulatory framework, there might be **joint investigations or co-regulation** of AI systems. Europe has some models: e.g., in the Netherlands, the data protection authority and financial authority have collaborated on fintech issues. The SCHUFA case will encourage more of that, because it highlighted both

[188] Ibid 61.

privacy and fairness angles. For future strategy, setting up **AI oversight task forces** that include multiple agencies could be beneficial for complex cases.

- *Global Influence and Convergence:* As mentioned, the logic of the SCHUFA decision might influence other jurisdictions with similar principles. The concept of requiring human option and contestability might find its way into more laws. In the U.S., there's already some movement: the FTC has recommended businesses provide human review for important automated decisions, and some proposed state laws echo GDPR-like rights. The lesson here is that regulators worldwide are learning from each other. If the EU shows it's enforcing these rights without economic collapse (banks haven't stopped lending, they'll adapt), others may be emboldened to adopt similar rules. On the flip side, regulators will watch for unintended consequences: Does requiring human oversight significantly slow down credit processes? Does it reduce the availability of credit to some because banks become more cautious? If there were evidence of that, regulators might need to fine-tune rules (ensuring they're not too onerous). So, another lesson is to **monitor the impact of regulation**. The EDPB and others might do follow-up studies on how the SCHUFA ruling's implementation affects industry and consumers, to inform future tweaks.

- *Empowering Individuals is Key:* Underlying GDPR and AAA is the idea of empowering those subject to algorithms. The case validated that concept. A single individual's persistence led to a change affecting millions. Future regulatory strategies might consider ways to lower the burden on individuals to achieve such outcomes. Perhaps via **collective redress** or regulatory interventions that don't wait for complaints. For instance, DPAs might conduct *own-motion investigations* of major AI systems rather than relying on a data subject complaint, which might take years to litigate. The EDPB could even identify certain AI use cases for proactive audit across the EU. For example, the EDPB could say: "We will scrutinise how credit scoring, recruitment algorithms, etc., are complying in 2025." This proactivity is part of moving from reactive to preventive regulation, which the AI Act and

AAA embrace. The lesson from SCHUFA is don't assume people will always go through lengthy court battles – regulators can step in earlier to examine these issues systematically.

- *Industry Adaptation and Innovation:* A positive takeaway is that regulation can spur **innovation for compliance**. Companies like SCHUFA might invest in better AI that is more explainable or in consumer-friendly portals for credit information. There's a burgeoning RegTech (regulation technology) sector and Fairness-Explainability-Transparency (FAT) tools for AI. Future strategies could encourage this, e.g., through regulatory sandboxes or pilot programs where companies test new, more transparent scoring methods under regulator guidance. The lesson here is that not all regulation is a cost; it can drive modernisation. Credit bureaus have been doing things roughly the same way for decades – now they have the impetus to update. If they innovate successfully (say, using AI to provide personalised advice to consumers on improving their score without revealing IP), that could be a win-win: consumers are helped, and companies differentiate themselves ethically.

In conclusion, the SCHUFA case emphasises aligning AI-powered decision systems with fundamental rights and human-centric values. It teaches that even entrenched industry practices must yield to the higher principle that **automation should serve people, not sideline them**. Regulators will carry this lesson forward, likely applying the same rationale to other automated decision domains, ensuring that the march of AI does not trample individual rights.

6.3 HireVue: AI-Driven Hiring Tools and Algorithmic Fairness

Case Overview: HireVue is a U.S.-based company offering AI-driven recruitment and hiring solutions. It became well-known for its **video interview platform**, where job candidates record themselves answering questions (often via webcam), and HireVue's AI algorithms analyse those recordings to provide scores or recommendations to employers. The system initially claimed to evaluate a candidate's **"employability" or fit** by examining not just the content of their answers,

but also their facial expressions, tone of voice, word choice, and other behavioural signals.[189] For example, HireVue's technology (in its earlier iterations) purported to measure traits like "cognitive ability" and "enthusiasm" by analysing micro-expressions or intonations.[190] Companies used these scores to decide whom to advance to the next hiring stage, in some cases automating the rejection of lower-scoring applicants. This promises efficiency – screening thousands of interviews quickly – but raises serious **legal and ethical questions** about bias, transparency, and validity in hiring. The HireVue case study explores how GDPR, the AI Act, and the AAA would apply to such AI in employment, and how regulators like DPAs, the U.S. FTC, and others have responded.

Legal Compliance under GDPR: When HireVue's tools are used by employers in the EU (or by EU-based subsidiaries of multinational companies), GDPR provisions become directly relevant. Several GDPR aspects come into play:

- *Lawful Basis & Data Minimisation:* The processing involved in AI video interviews is extensive. HireVue's system collects video (imagery of the candidate), audio, transcribes speech to text, and may derive various inferences (like sentiment or perceived personality traits). This encompasses personal data and potentially **special category data**. For example, a video inherently reveals someone's **race/ethnicity** and **gender**, and possibly other sensitive traits (like if the person has a disability affecting speech or movements). Under GDPR, processing data that even indirectly reveals such characteristics (biometric data for identification, health/disability status, etc.) requires a lawful basis under Article 6 and potentially an Article 9 exception if considered special category data. Employers typically cannot easily claim a legal obligation or vital interest for this, so they often seek **consent**. Indeed, some implementations of HireVue in Europe reportedly asked candidates to consent to the video processing. However, consent in an employment context is legally fraught – GDPR Recital 43 notes it is unlikely to be "freely given" due to the power imbalance

[189] *HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition* (Electronic Privacy Information Center, 12 January 2021) https://www.epic.org/hirevue-facing-ftc-complaint-from-epic-halts-use-of-facial-recognition/
[190] Ibid.

(candidates may fear they won't be considered if they refuse). Thus, reliance on consent could be challenged as invalid. Alternatively, employers might argue **legitimate interests** (Art. 6(1)(f)) in using innovative hiring tools to improve efficiency. But that requires a balancing test: the employer's interest vs. the candidates' privacy and data protection rights. Given the intrusiveness and novelty of AI analysis of faces and voices, a DPA or court might find the candidates' rights prevail, unless strong safeguards exist. Additionally, if special category data are effectively being processed (e.g., biometric data if the system does any facial recognition or analysis that uniquely identifies aspects of a person's face), an Article 9 condition like "explicit consent" would be needed. In short, **the lawfulness of HireVue under GDPR is at best debatable** without explicit consent, and consent may be deemed not freely given. One could conceive a candidate complaining that they were forced to undergo an AI evaluation or else lose a job opportunity, a scenario ripe for GDPR enforcement, given the power imbalance.

- *Article 22 – Automated Decision-Making:* Much like credit decisions, hiring decisions can fall under GDPR's Article 22. If HireVue's AI is used to **solely automate** rejection of candidates (e.g., automatically filtering out those below a certain score with no human vetting), that is a decision producing a significant effect (not getting a job is clearly significant for an individual). GDPR would **prohibit** such practice by default, unless an exception applies. Unlike credit scoring, hiring isn't clearly "necessary for a contract" (some might argue selecting an employee is necessary to fulfil an employment contract, but that's a stretch and not what Art . 22 had in mind). There's typically no law authorising AI-only hiring. Consent could be an option, asking the candidate, "Do you consent to be evaluated only by AI?" but again, consent under duress of wanting the job is problematic. Therefore, to comply, most deployments of HireVue in Europe ensure a **human is in the loop**. Companies often describe the tool as assisting recruiters, not replacing them. If challenged, they'd likely assert that a human recruiter reviews the AI's recommendations, thus it's not

"solely automated." However, the reality might be that overwhelmed recruiters indeed rely heavily on the AI's rank-ordering, meaning the decision is *effectively* automated. The recent SCHUFA reasoning could be extended here: if the AI's score "guides" decisions about hiring similarly significantly,[191] then Article 22 would apply. Employers would then need to offer the candidate the safeguards, e.g., inform them they can request a human re-evaluation.

Even aside from outright automation, GDPR requires **transparency** (Articles 13 and 14) about the use of such AI. Candidates should be informed that AI will analyse their video, the purposes of analysis, and some logic of how it works (at least the features considered). The **right to explanation** under GDPR isn't explicit but is derived from the transparency and access rights combined with Article 22. In an HR context, one could analogise that a candidate rejected due to an AI assessment can ask for an explanation of that decision. Indeed, several DPAs (like France's CNIL) have indicated that candidates should be able to get the reasons, even if it's just a high-level explanation. HireVue's complex algorithms might make that challenging, but legally, the employer (as the data controller) would have to supply meaningful information. If they simply say, "the algorithm scored you low," that might be inadequate under GDPR, especially after the CJEU's broad stance in SCHUFA that even intermediate algorithmic outputs count.

Data Protection Authorities have started paying attention to AI in hiring. For instance, the **Spanish DPA (AEPD)** issued guidance on AI in selection, warning against opaque algorithms and reminding employers of Art. 22. The **Italian DPA (Garante)** in 2021 went so far as to ban a system that purported to profile employees' personalities via AI (a different scenario, but it shows their approach). The Garante stated that such processing was unlawful without a proper legal basis and violated principles of accuracy and fairness. HireVue's system, especially in its earlier form with facial analysis, might have faced a similar fate: a DPA could call it **excessive and not sufficiently justified**. The **UK's ICO** has also been vocal: it released an "AI and data protection risk toolkit"

[191] Almada (n 83) 54.

including hiring scenarios, emphasising data minimisation (only collecting data relevant to job performance) and algorithmic fairness.

In terms of special GDPR provisions: If HireVue's analysis is deemed to be making **psychological inferences** about a person (like cognitive ability or personality traits), that might even raise questions under GDPR's provisions about profiling and possibly even medical data (if, for instance, it inadvertently profiles mental health or neurological patterns). These are grey areas, but they show potential vulnerabilities under GDPR.

Compliance under the EU AI Act: Under the AI Act, hiring tools like HireVue's are explicitly designated as **high-risk AI systems**. Annexe III of the Act includes AI systems intended for "**employment, workers management and access to self-employment**," particularly for recruitment or selection (CV sorting, interviewing, aptitude testing).[192] This means HireVue (or any similar system) would have to comply with a suite of obligations:

- **Data and Bias Risk Management:** The provider (HireVue) must implement a risk management system (Article 9) to continually assess and mitigate risks to fundamental rights. Concretely, for a hiring AI, this would focus on **bias mitigation**. HireVue would need to evaluate its algorithms for bias against protected classes (gender, ethnicity, age, etc.) and adjust them or the data to minimize any disparate impact. It would also need to ensure that the features used have a logical link to job performance (to avoid arbitrary or prejudice-laden criteria). Technical documentation (Article 11) should detail how the model was trained and tested for bias. For example, if the AI uses voice tone, HireVue would have to document why that is relevant and how it avoids penalizing, say, non-native speakers or people with speech disabilities.

- **Transparency and User Information:** Article 13 of the AI Act will require that providers of high-risk AI give clear information to the users (in this case, the hiring company, and by

[192] Ibid. 60.

extension some information should be passed to candidates) about how the system should be used, its limitations, and what it is (and isn't) evaluating. HireVue would likely have to supply **instructions** to employers that, for instance, the tool should not be used as the sole criterion, that it may have certain error rates, and that employers should inform candidates they are being evaluated by AI. Additionally, the AI Act has provisions (Article 52 in some versions) that high-risk AI interacting with people must notify those people that they are interacting with AI. In hiring, that translates to **candidates being told clearly that an AI is involved in assessing their interview** (which overlaps with GDPR's transparency requirement, reinforcing it).

- **Human Oversight Measures:** The AI Act requires that high-risk AI be designed so that **effective human oversight** is possible (Article 14). For HireVue, that might mean the system should allow a human recruiter to easily review the candidate's original video and the AI's analysis, to ensure no obvious mistakes. It might also mean building in controls to prevent automation bias e.g., the interface could encourage human reviewers to use their judgment and not blindly accept the AI ranking (perhaps by not presenting the AI score as absolute or by grouping candidates into bands rather than precise ranks, etc.). The Act, in spirit, wants to avoid "automation bias" by design.[193] HireVue would need to demonstrate that its product doesn't lure users into over-reliance. This could be done by appropriate training for users or by product design (like always requiring a human to confirm rejections).

- **Accuracy and Robustness:** Under Article 15, HireVue would have to ensure the system's accuracy is appropriate for its purpose and is robust against errors or manipulation. This addresses a concern: AI in hiring could be fooled or might perform poorly in unexpected conditions (like if a candidate's video or audio quality is bad, does it unfairly lower their score?). Ensuring robustness might mean setting minimum video quality requirements or

[193] Ibid 61.

detecting when the system is not confident so it can flag those cases for human handling rather than making a possibly erroneous evaluation.

- **Conformity Assessment:** HireVue or its distributor in Europe would need to undergo a conformity assessment (likely a self-assessment if using harmonized standards, or involving notified bodies if not) and affix a CE mark. This involves preparing a comprehensive technical file with all the documentation above, a detailed risk assessment, and an EU declaration of conformity. It's a bureaucratic but important step essentially verifying that from design to testing, the product meets the AI Act's criteria. If HireVue can't show that (say it can't sufficiently mitigate bias or lacks transparency), it would not be allowed to be sold/deployed in the EU.

The AI Act thus ensures that **before** such a tool is used at scale in the EU, it meets certain quality and fairness benchmarks. Had these been in place earlier, some of HireVue's controversial features (like facial expression analysis) might have been flagged or disallowed unless validated. Indeed, emotion recognition in employment is going to be banned by the AI Act (Article 5(1)(f) use of AI for emotion evaluation in hiring is prohibited without exceptions[194]). HireVue's earlier claims of analyzing "emotional intelligence" via facial cues would likely violate that outright ban. The company preemptively removed those features by 2021 (likely anticipating such pushback).[195] Under the AI Act, continuing to use emotion analysis in hiring would be illegal, period.

Compliance under the Algorithmic Accountability Act (AAA): The Algorithmic Accountability Act would directly cover AI hiring tools as part of "automated decision systems" impacting employment opportunities.[196] For an AI like HireVue's, the AAA would require two main things: **impact assessments** and **ongoing oversight/reporting**.

[194] Ibid.
[195] *HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition* (n 187).
[196] Algorithmic Accountability Act of 2023: Summary (n 142).

- **Algorithmic Impact Assessment (AIA):** The employer deploying HireVue, or HireVue itself if providing the service, would need to conduct an assessment focusing on risks of bias, discrimination, privacy invasion, and overall validity.[197] For instance, they'd have to analyze: Does the AI systematically score certain groups of candidates lower? What data is it using is any of it irrelevant or sensitive (like analyzing facial structure which could tie to ethnicity)? Are there any less discriminatory alternatives? The AAA also specifically mentions examining whether personal attributes like race, sex, or other sensitive traits influence outcomes[198] essentially a bias audit. If issues are found, the entity must devise mitigation measures. For HireVue, a mitigation might be removing or down-weighting features that correlated with, say, gender in a way unrelated to job performance.

- **Transparency and Reporting:** AAA would likely require that companies inform individuals when they are subject to automated assessments. In a hiring scenario, this translates to candidates receiving a notice: "Your interview will be analyzed by an AI system." AAA also pushes for **public disclosure** of the existence of such systems. Possibly, larger companies might have to report to the FTC that they are using AI in hiring and provide summary results of their bias audits.[199] The FTC could publish anonymized aggregate trends (AAA mandates an annual report by FTC) so the public and policymakers know how prevalent these practices are and what issues are being found.

- **FTC Enforcement:** If a company like HireVue failed to do an adequate assessment or ignored obvious bias, the FTC could take action under AAA regulations (once promulgated). That might include fines or orders to stop using the AI until fixed. The FTC already has authority it can use (if an AI tool is "deceptive" or "unfair"), for example, if HireVue advertised its tool as bias-free and objective but it wasn't, that could be a deceptive claim. EPIC's 2019 complaint to the FTC about HireVue alleged exactly that: that HireVue's

[197] Ibid.
[198] Ibid.
[199] Ibid.

claims were deceptive and the practice was unfair to consumers (job applicants).[200] While AAA isn't law yet, the FTC is signaling that algorithmic fairness is within its remit. The AAA would solidify that and give more structure (via required assessments).

- **Overlap with Employment Law:** AAA would sit alongside existing civil rights laws like Title VII of the Civil Rights Act (which prohibits employment discrimination). The EEOC (Equal Employment Opportunity Commission) is the main enforcer of those laws for hiring discrimination. In an interesting development, the EEOC and DOJ have interpreted that using AI that disproportionately screens out protected groups could violate the law, and employers cannot evade liability by saying "the algorithm did it." They issued guidance on how AI might violate the Americans with Disabilities Act if it doesn't accommodate people with disabilities (for example, a video interview AI might disadvantage someone with autism or speech impairments, which could be disability discrimination unless accommodations or exceptions are provided). So even outside AAA, U.S. regulators are active: the EEOC can investigate a company if an AI tool is found to systematically discriminate. The AAA would complement this by requiring proactive checking for such discrimination rather than waiting for a complaint.

So, AAA's adequacy for HireVue-like systems lies in **forcing self-scrutiny and documentation of fairness efforts**, which is currently largely voluntary. Some progressive companies do internal AI fairness audits, but AAA would make it a standard practice across the board.

Ethical Implications: Fairness and Bias: AI in hiring crystallises concerns about **algorithmic bias** and fairness. Ethically, hiring should be based on merit and job-relevant qualifications, not on attributes like race, gender, or other irrelevant traits. Yet, AI systems trained on past hiring data can inadvertently learn biases for example, if a company historically hired mostly men, an AI might pick up on subtle patterns that correlate with male candidates and score female candidates lower (as happened in Amazon's famous AI recruiting tool that was abandoned for penalising resumes with

[200] *HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition* (n 187).

indicators of being female[201]). There's also the risk of **cultural or ableist bias**: HireVue's analysis of tone, facial expressions, or word choice might favour neurotypical, extroverted, or culturally "mainstream" styles of communication, disadvantaging those who communicate differently but could still be excellent employees. For instance, a person on the autism spectrum might speak in a flat tone or avoid eye contact, and an AI not attuned to diversity could rate that as low "enthusiasm" or "sociability," unfairly labelling the candidate as a poor fit. This raises ethical red flags about **discrimination against neurodiversity or introverts**. Similarly, candidates who speak English with an accent or non-native fluency might be scored lower on communication by AI, even if the job doesn't actually require perfect English, raising issues of ethnic and linguistic bias. Ethically, the use of AI in hiring must strive for **procedural fairness** (treating like cases alike and different cases differently, based on relevant factors only) and **distributive fairness** (not systematically excluding certain groups without justification). There's also the principle of **equal opportunity**: AI should not become a gatekeeper that reinforces historical inequalities or denies opportunities to qualified individuals due to irrelevant correlations.

- **Transparency and Consent:** Ethically, candidates should know they are being evaluated by an algorithm and ideally have some understanding of what traits or criteria are being assessed. Transparency in hiring AI is tied to **respect for persons,** treating candidates not merely as data points but as individuals with a right to information about processes affecting them. A lack of transparency in a black box hiring decision can leave rejected candidates confused and suspicious (e.g., "Was I rejected because of an AI's flawed perception of my facial expression, rather than my actual ability?"). This opacity can also undermine trust in the employer and in the fairness of the hiring process at large. As for consent, while in practice candidates often have little choice, ethically it's problematic to subject someone to analysis of their facial movements or tone without explicit, informed consent. If a candidate isn't comfortable with an AI judging them, should they have an alternative (like requesting a

[201] Max Langenkamp, Allan Costa and Chris Cheung, *Hiring Fairly in the Age of Algorithms* (2020) arXiv, s 2.1 DOI: 10.48550/arXiv.2004.07132 https://doi.org/10.48550/arXiv.2004.07132

human interview)? Some ethicists would say yes, providing an opt-out respects individual autonomy, though in reality, opting out might disadvantage them if the company heavily relies on AI. This is a genuine ethical dilemma: requiring an alternative path to avoid coercion, versus the company's interest in using a standardised process. At a minimum, **informed consent or notice** is ethically necessary: candidates shouldn't be tricked into thinking a human is reviewing their interview when it's actually AI.

- **Accuracy and Validity:** There's an ethical obligation for any assessment tool to be **accurate and job-related**. Early claims by HireVue that it could infer traits like cognitive ability or "professional demeanour" from video analysis were met with scepticism by psychologists, raising concerns that decisions might be made on **junk science**. Using an invalid or pseudoscientific method to judge candidates is unethical because it can arbitrarily harm individuals' career prospects. If, for example, the AI assigns a low score due to a candidate's accent or camera shyness (which have no bearing on job performance in many roles), that is an **unjust harm**. Ethically, the use of AI should be governed by the principle of **non-maleficence** (do no harm) and **beneficence** (do good). If the AI's benefit (efficient screening) is outweighed by harm (good candidates being unfairly filtered out, or diverse candidates being disproportionately rejected), that's ethically unacceptable. Thus, there's an ethical impetus to thoroughly validate AI tools in hiring to demonstrate that their criteria correlate with actual future job performance and do not systematically misjudge certain groups. If that validation can't be shown, ethically, the tool shouldn't be used.

- **Human Dignity and Agency:** A core ethical concern is that AI hiring tools might reduce candidates to a set of algorithmic scores, undermining the **human dignity** of the applicants. The experience of being evaluated by a machine can be dehumanising, especially if one suspects the machine might be making shallow judgments (like analysing your facial expression rather than your actual skills). Additionally, if an AI rejects someone, the candidate often doesn't have a chance to advocate for themselves or explain context, which

removes human dialogue and mercy from the equation. For example, a candidate might have a justified reason for a nervous demeanour (maybe they are first-generation college grads unaccustomed to interviews), a human interviewer might see potential beyond nervousness, whereas an AI might not. Ethically, many argue there should remain **human agency in decisions that deeply affect people's lives**, like getting a job. Completely delegating that to AI could be seen as an abdication of moral responsibility; hiring involves moral judgments (e.g., giving someone a chance) that some believe shouldn't be made solely by algorithms.

- **Accountability:** The ethical principle of accountability demands that when a decision is made, it's clear who (or what) is responsible and that the decision-making process can be scrutinised and, if necessary, corrected. AI complicates this; the line of accountability can blur between the tool provider and the employer user. A candidate rejected due to HireVue might not know whether to blame the employer, the algorithm, or their own performance. From a justice perspective, that's problematic. Ethically, employers should remain accountable for their hiring decisions, even if aided by AI. And they should be prepared to answer a candidate who asks, "Why was I not selected?" with more than, "The computer said no." The concept of **due process** comes in: just like in SCHUFA, people have a right to contest. In hiring, there's an ethical argument (though not a strong legal right) that candidates should be able to get feedback or contest an unfair decision. Of course, historically, employers haven't been legally required to give rejected candidates an explanation (except in some contexts like internal promotions or public sector jobs), but the introduction of AI might shift expectations ethically if not legally. There's a stronger case that algorithms' decisions should be reviewable to ensure they weren't based on flawed logic.

- **Beneficence (Potential Upsides):** From an ethical angle, it's also worth noting the potential positive ethical implications if done right: AI could help reduce human biases (like nepotism or first impression bias) if properly calibrated. For example, HireVue claims it removed

facial analysis partly to focus on more relevant, less bias-prone metrics, and that it audits its algorithms for fairness. If an AI hiring tool can be proven to *increase* fairness, say, by focusing on competencies in a standardised way and ignoring demographic factors, then ethically it might be justifiable or even laudable. Some ethicists argue that a well-designed AI might actually be less biased than a human manager who might subconsciously favour candidates of their own gender or background. However, that's a big "if" and requires rigorous evidence. So far, scepticism remains, but ethically, one should consider the nuance: The **goal** is not to demonise AI but to ensure it is used in a way that genuinely **promotes fairness, transparency, and respect**.

Role of Regulators: Data Protection Authorities (EU DPAs): EU DPAs have jurisdiction when personal data is processed in hiring via AI. While no DPA has publicly sanctioned an employer for using HireVue specifically (as of 2025), they have shown increasing interest in the topic. For instance, France's CNIL published guidance on AI and discrimination, and mentioned hiring tools as an area to watch. The **EDPB** included "algorithmic transparency" in its work program and might develop guidelines for AI in HR. If a candidate in the EU lodged a complaint that they were unfairly evaluated by an AI tool, a DPA could investigate the employer for GDPR violations (e.g., lack of valid consent or Art. 22 breach). It's plausible that as awareness grows, DPAs will receive such complaints. Alternatively, DPAs might proactively conduct sectoral investigations – for example, checking how large companies recruit and whether they comply with GDPR in doing so. The Spanish DPA's guide on AI in recruitment suggests they are ready to enforce if needed.

One interesting angle: some EU countries (like Austria and Germany) have co-determination rules requiring informing or consulting employee representatives (works councils) when introducing algorithmic management tools. HireVue for hiring might not strictly fall under those, since candidates aren't employees yet, but it indicates a broader labour law intersection. DPAs might coordinate with labour inspectors if needed.

- **Equal Employment Regulators (EU):** In the EU, anti-discrimination in hiring is also regulated by equality bodies (like the Equality and Human Rights Commission in the UK, or similar agencies in EU countries). They can handle complaints of discrimination. If an AI system systematically disadvantages, say, women or minorities, that can lead to legal liability under employment discrimination laws. However, proving that can be challenging without transparency. Regulators might need to demand the algorithm's data or call for expert analysis. In a forward-looking move, some equality regulators might issue guidance to employers: e.g., the Dutch Human Rights Institute might advise employers using AI to ensure no prohibited bias. Coordination between DPAs and equality bodies could be key – e.g., a DPA can enforce the transparency so that equality bodies can inspect for bias.

- **Federal Trade Commission (FTC, U.S.):** The FTC has asserted authority over AI biases as potentially "unfair practices." In April 2021, the FTC warned businesses that selling or using racially biased algorithms could violate the FTC Act Section 5 (unfair practices). The EPIC complaint against HireVue was essentially a call for the FTC to act.[202] While the FTC did not publicly fine HireVue, the pressure arguably contributed to HireVue dropping its most contentious feature (facial analysis).[203] The FTC also endorsed the idea of algorithmic transparency in its 2022 report to Congress. If AAA passes, the FTC would take on a bigger role, but even without it, the FTC might pursue a particularly egregious case if one arises (e.g., if evidence emerged of a hiring AI that was blatantly biased or if a vendor lied about its tool's capabilities).

- **Equal Employment Opportunity Commission (EEOC, U.S.):** The EEOC has been actively monitoring AI in hiring. In 2023, it resolved some of its first discrimination cases involving AI for example, one employer's AI screener that inadvertently disqualified people with certain disabilities led to an EEOC settlement requiring accommodations. The EEOC

[202] *HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition* (n 187).
[203] Ibid.

also held public hearings on AI and issued technical assistance documents (like how AI could violate the ADA). They have set up an "AI and Algorithmic Fairness Initiative." One can foresee the EEOC bringing a landmark case to establish principles, such as suing an employer for using an AI hiring tool that has a disparate impact on a protected group. The remedy could involve requiring the employer to stop using the tool or to modify it. The EEOC can subpoena data, so they might compel a company or vendor to provide the training data and results to analyse bias.

- **State and Local Regulators (U.S.):** Some jurisdictions have stepped in: New York City's law on Automated Employment Decision Tools (effective 2023) requires that employers using such tools conduct a **bias audit** annually and publish a summary, as well as notify candidates about the tool and allow them to request an alternative process. The NYC Department of Consumer and Worker Protection is enforcing this (with fines for non-compliance). This is effectively a regulation akin to AAA, but at the city level. Illinois' Video Interview Act requires notifying candidates and sharing their scores upon request, and was recently amended to require an annual report of aggregate demographics of who was hired vs. not, to watch for biases. These localised laws mean regulators on the ground (city agencies, state attorneys general) might investigate companies. For example, if a company in NYC uses HireVue without doing a bias audit, it could face penalties. These state/local rules often influence others. California is considering similar measures, and so on.

- **Self-Regulation and Industry:** Apart from government regulators, industry standards bodies and certification groups have started emerging. For instance, the IEEE and ISO are working on standards for algorithmic bias and quality. Companies like HireVue often tout that they are audited by third parties or certified for bias mitigation. There are independent auditors (like ORCAA or Verité) that examine AI hiring tools. So a form of quasi-regulation is developing through certifications (the way ISO 9001 certifies quality processes, we might see "fair AI" certifications). Regulators in the future might incorporate these, maybe giving

credit to companies that use certified tools or even requiring a certification as part of compliance.

- **Courts and Litigation:** Regulators aside, the threat of litigation (especially in the U.S.) is another pressure. A rejected candidate might sue an employer alleging the AI had a disparate impact on, say, older applicants (age discrimination). While it's tough to get data to prove it, even the discovery process could force the employer to reveal information about the algorithm. The possibility of class actions (a group of candidates who believe the AI was biased) could prompt employers to be cautious. To preempt this, regulators could encourage companies to collect and analyse data on how different groups fare in AI assessments (which NYC now mandates). If a pattern of discrimination is found, it must be justified by business necessity and lack of less discriminatory alternatives a high bar. The ethical and legal consensus seems to be forming that simply saying "the algorithm is a trade secret" will not shield companies from scrutiny when rights are at stake.

Practical Outcomes or Enforcement Actions: As noted, one major outcome so far was **HireVue discontinuing its facial analysis feature in 2021**.[204] This change was directly attributed to public and regulatory pressure (EPIC's FTC complaint, media scrutiny, and general outcry from AI ethicists). HireVue acknowledged that the feature wasn't worth the controversy, which is both a compliance and PR decision. Practically, this may have reduced some of the more egregious bias risks (since facial expression analysis is scientifically dubious and prone to bias). HireVue now focuses on analyzing speech and perhaps cognitive game results (they also acquired a game-based assessment company). It shows that when regulators and civil society raise alarms, AI vendors can pivot to mitigate perceived risks.

- The use of AI in hiring has continued to grow, but with more caution and **attempts at fairness controls**. Many vendors now openly talk about bias testing and employing industrial-organisational psychologists to ensure validity. For example, HireVue has claimed

[204] Ibid.

they audit algorithms for bias (e.g., ensuring selection rates for different demographic groups are within acceptable bounds, and if not, adjusting the model). Some vendors even allow an external observer, like a professor, to review their algorithms under NDA. These steps, while perhaps partly marketing, indicate that the **norms are shifting;** what was the wild west is becoming more regulated by expectation if not law yet.

- **Candidate Reactions and Adaptations:** A practical development is that candidates are becoming aware of AI interviews, and some have tried to "game" them. There are now articles advising applicants on how to perform in an AI interview (e.g., speak clearly, use keywords from the job description, maintain eye contact with the camera, etc.). Some even suggested exaggerating facial expressions or enthusiasm to score better (though with HireVue dropping facial cues, that's less relevant now). This is reminiscent of how people adapted to standardised tests, they learn test-taking strategies. Ethically, this is interesting: if people try to "beat" the AI, it could reduce the tool's usefulness (everyone might start sounding the same or overly polished, making it hard to differentiate true skill). As a result, some employers might see diminished returns from AI interviews and could scale back usage.

- **Enforcement examples:** In New York City, initial enforcement of the bias audit law has been rocky many employers weren't ready by the January 2023 start date, causing delays in enforcement. But by mid-2023, companies began posting biased audit summaries. These revealed some insights (like Tool X had almost no disparity between genders but some disparities in race, etc.). If those audits had shown extreme bias, presumably, regulators would step in to ban or require changes. So far, audits are showing some disparities, but often within legal thresholds (e.g., selection rate for a group is at least 80% of the highest group's rate, the "4/5ths rule"). If an audit fails, the employer either must stop using the tool or retrain it and re-audit. This shows a model of how enforcement can drive remediation quietly.

Illinois, under its law, collected annual reports on video interview usage. The 2020–21 report showed relatively low adoption and no clear demographic biases, but it's a small sample. These transparency measures are practical enforcement tools because they allow the regulators and public to see patterns and push for improvement without immediately resorting to lawsuits.

- We have yet to see a blockbuster lawsuit or fine purely about an AI hiring tool's bias. But I suspect one is on the horizon. Possibly a case where an employer inadvertently screened out people with disabilities (like an AI that relies on speed of response might disadvantage someone with a speech impediment or who uses assistive tech, a potential ADA violation). The DOJ/EEOC have warned about that: for example, an AI that automatically rejects anyone who doesn't score above X on a timed game could be illegal if it doesn't accommodate those who might need more time due to a disability. Enforcement in this area might first come via such disability discrimination cases because they're a bit more straightforward (the law mandates reasonable accommodation, so if AI doesn't allow for that, it's a direct violation).

One notable enforcement-like action: the California Department of Fair Employment and Housing (DFEH) reached a settlement in 2021 with a company whose online hiring exams (not exactly AI, but digital assessments) had accessibility issues requiring them to accommodate and consider alternative methods for disabled applicants. That logic extends to AI interviews.

- Another outcome: Some employers, fearing these issues, have **stepped back from AI** for certain roles. There have been reports of companies trialling HireVue or similar, then discontinuing due to mixed results or negative candidate feedback. The labour market also influences this; in a tight labour market, companies might avoid impersonal AI processes so as not to deter scarce candidates. In a looser market, they might use it more. So economic context can drive usage more than regulation in some periods.

- **Market impact:** The scrutiny on HireVue and others has also led to new startups focusing on more transparent or "ethical" AI hiring tools, or tools that only do simpler tasks (like scheduling or basic skill screening) to avoid bias minefields. Traditional assessment companies (like those that make cognitive or personality tests) are blending AI, but doing so carefully under existing testing standards (like the US EEOC's Uniform Guidelines on Employee Selection, which apply to any selection device, AI or not – requiring validation studies). So practically, we might see a convergence where AI hiring tools start looking more like traditional assessments (with documented validity) plus some machine learning enhancements, rather than inscrutable models.

Lessons Learned & Implications for Future Regulation: *Early Self-Correction vs. Waiting for Law:* The HireVue case shows that sometimes the industry will adjust before formal regulation compels it, especially under combined pressure from advocacy and potential regulator intervention. HireVue dropping facial analysis is a key example.[205] It indicates a lesson: **if certain AI practices are broadly seen as unethical or legally risky, proactive change can forestall heavier-handed regulation**. For future regulatory strategy, this suggests regulators can use a mix of tools – not just fines, but also **guidance and public pressure**. By articulating clear expectations (like "emotion AI in hiring is not acceptable"), regulators can prompt companies to abandon high-risk features on their own.

- *Need for Standards and Audits:* A recurring theme is the need for consistent **standards** in evaluating AI fairness. In hiring, what metrics define "bias"? The 4/5ths rule? Significance testing? What sample size? The NYC law forced the question, and we saw audit methodologies emerging. Future regulation should incorporate or reference **technical standards** for algorithmic audits. The lesson is that without standardised methods, it's hard to compare or enforce claims of fairness. Efforts like the **EEOC's draft guidance on algorithmic fairness** or the **IEEE's standards** will be crucial. Regulators might even

[205] Ibid.

certify third-party auditors or tools for bias detection. This could evolve similarly to financial auditing – maybe large AI providers will have to get an independent "bias audit" opinion annually. The AAA explicitly leans this way (structured assessment under FTC oversight). The EU AI Act also implicitly expects providers to document risk mitigation. So the implication is that **systematic auditing will become a norm**, and regulation should ensure those audits are rigorous and not mere check-the-box exercises.

- *Human-Centric Approach Preservation:* Much like with credit scoring, regulators will likely insist that **AI should assist, not replace, human judgment** in areas like hiring. The lesson from HireVue's controversies is that removing humans entirely can lead to backlash. Future regulatory frameworks (or even company policies shaped by regulation) might encode that certain decisions always need a human confirmation, especially negative ones. For instance, one could imagine a rule: "No candidate shall be rejected by AI alone; a human must review borderline cases or all cases flagged by the AI as rejected." This is somewhat already law in GDPR's Article 22, but might be further reinforced by labour laws or AI Act guidance for high-risk systems. The broader point is a **socio-ethical stance becoming regulatory: keep humans meaningfully involved to uphold accountability and empathy**. The White House's 2022 Blueprint for an AI Bill of Rights included a principle that individuals should have access to a human alternative, which might inspire future regulations.

- *Cross-Disciplinary Training and Oversight:* Hiring AI implicates data protection, anti-discrimination, and industrial-organisational psychology (validity of tests). A single regulator often doesn't have all the expertise. The lesson here is that governance might require **multi-disciplinary teams**. For example, a DPA might bring in an I/O psychologist to assess if an AI's assessment correlates to job requirements (to gauge fairness beyond pure data issues). Or an EEOC investigator might coordinate with the employer's data scientists to understand the algorithm. To facilitate this, regulations might call for documentation not just of bias metrics but of validity evidence (like, did the AI's scoring align with later job

performance of hired candidates?). If an AI is proven ineffective, continuing to use it could be considered unfair to candidates. So future strategies might incorporate **effectiveness audits** too, not just bias – ensuring AI is actually doing what it claims (these ties to consumer protection against snake oil AI).

- *Candidate Rights and Experience:* We might see moves to strengthen candidate rights in automated processes. Perhaps requiring that candidates can request an **alternative assessment method** if the AI might disadvantage them (similar to disability accommodation). For instance, a person with a speech impediment might request a written Q&A instead of a video interview. Ethically and under ADA, that should be offered. Regulators like EEOC encourage that. Thus, a best practice (and possibly future rule) is to always offer a human interview or other format if requested. This aligns with fairness and could even be framed under GDPR's right to contest automated decisions effectively, "I contest the AI's assessment, I want a human interview." The lesson is empowering individuals to not be locked into one algorithmic fate. If such rights become standard (e.g., mandated in EU by an update or explicitly in national law), it ensures AI doesn't become an inflexible gate.

- *Global Norms Setting:* The combination of EU's strong data protection, US's emerging algorithmic accountability, and local laws like NYC's is creating a patchwork that could converge into global norms. Multinational companies may apply the highest standard across all operations for simplicity (e.g., if a company must do bias audits for NYC and ensure GDPR compliance in EU, they might just do it everywhere). The lesson here is that **leading regulations can uplift global practice**. We see HireVue itself applying changes globally (they dropped facial analysis not just in EU but worldwide). Similarly, if AAA or EU AI Act pushes things like documenting bias mitigation, big vendors will do that uniformly. So even if not all countries regulate AI hiring, the practical effect might spread. Regulators should be

mindful of this leverage by setting high standards, they can indirectly protect people beyond their jurisdiction.

- *Continuous Monitoring:* AI models are not static – they evolve (retraining on new data, etc.) and their impacts can shift as user behavior changes (candidates might adapt, as mentioned). So one-off audits may miss issues that crop up later. Therefore, regulators are likely to require or encourage **continuous monitoring** of AI performance. AAA's reporting obligations and the AI Act's post-market monitoring reflect that.[206] The lesson is treat AI management like a continuous quality assurance process, not a one-time certification. Employers using AI should periodically check outcomes and calibrate. A future regulatory idea could be requiring large employers to publish annual fairness reports of their hiring (not unlike diversity reports some publish voluntarily). This transparency would push companies to improve year over year.

- *Public Awareness and Education:* Another lesson is the importance of public awareness. Many candidates initially didn't realize a machine was reading their videos. Now, with press coverage and laws requiring notice, they know. This awareness can create a form of public accountability – companies might not want to be known as "the one that rejects humans by AI with no mercy." As AI gets regulated, efforts to educate affected people about their rights and how the AI works (to the extent possible) should be part of the strategy. Regulators might publish plain-language resources or require that companies do so as part of transparency (like NYC requires posting a summary of how the tool works). In essence, **empowering the public** is a regulatory tool in itself.

In conclusion, AI-driven hiring tools like HireVue's present a microcosm of the broader challenge of governing AI: balancing innovation and efficiency with fairness, accountability, and human rights. The steps taken so far, removal of problematic features, introduction of bias audits, and

[206] Algorithmic Accountability Act of 2023: Summary (n 142).

nascent laws indicate an evolving consensus that such AI must be kept on a tight ethical and legal leash. The adequacy of GDPR and AAA in this context will be seen in how well they enforce those principles. GDPR provides individual rights and principles, the AI Act will impose design-time obligations, and AAA would bring systematic oversight together; they form a lattice of protection. The key lesson is that **multi-layered regulation, involving technical standards, ethical principles, and legal mandates, is necessary to ensure AI in hiring serves as a tool for good rather than a vector of unseen discrimination or unfairness**.

Conclusion:

Across these case studies, Clearview AI's mass surveillance tool, SCHUFA's automated credit scoring, and HireVue's AI hiring system, we see common threads and crucial differences in how current laws address AI-powered decision-making. The **GDPR** has proven to be a powerful instrument for protecting fundamental rights against AI-related harms: it was used to declare Clearview's biometric dragnet illegal, and (with help from the CJEU) to bring credit scoring and hiring algorithms under the ambit of data protection principles like transparency, lawfulness, and fairness. However, GDPR enforcement has also shown limits. It is largely **reactive**, intervening after harms or complaints, and faces challenges with extraterritorial actors or proving algorithmic bias within its primarily privacy-focused provisions.

The **EU AI Act** is set to bolster this governance by adding an **ex ante regulatory layer** targeted specifically at AI. It takes into account context and risk (e.g., treating employment AI or credit AI as "high-risk" requiring compliance before deployment) and addresses facets that GDPR does not explicitly cover, such as technical robustness, accuracy, and the intricacies of algorithmic bias mitigation. The synergy of GDPR and the AI Act will create a comprehensive framework: **GDPR ensuring rights and general principles**, and the **AI Act ensuring rigorous development and deployment standards**. As seen in each case, many of the issues (biometric surveillance, biased scoring, opaque hiring algorithms) are as much about fundamental rights and ethics as about personal data. The AI Act broadens the regulatory gaze to **fundamental rights at large** (e.g.,

non-discrimination, human oversight, societal risks), which is exactly where GDPR's mandate might stop. Together, they aim to ensure not only that AI doesn't violate data protection, but that it doesn't undermine human dignity, equality, or safety.

In the United States, the **Algorithmic Accountability Act (AAA)**, though not yet law, encapsulates a growing recognition that **algorithmic processes need accountability and oversight akin to financial auditing or environmental regulation**. The AAA's approach requiring impact assessments and transparency for AI systems in critical domains reflects lessons learned from cases like these: that companies should not be trusted to self-police algorithms without external checks, and that consumers (whether citizens, borrowers, or job-seekers) deserve protections from hidden, potentially prejudiced decision-making. The AAA would fill gaps in U.S. law where, currently, only a patchwork of sectoral laws (like credit and employment discrimination statutes) and general consumer protection principles apply. The case studies underscore how, in the absence of something like AAA, enforcement in the U.S. relies on creative uses of old laws (e.g., Illinois' BIPA for Clearview, Title VII for hiring algorithms, or FCRA for credit models), tools that can address symptoms but not always the algorithmic root cause.

Each case also brings out the role of **regulatory bodies** and the need for them to evolve and cooperate. Data Protection Authorities in the EU have been key enforcers, but they will need to collaborate with new AI Act authorities, consumer protection agencies, and sector regulators. In the U.S., agencies like the FTC and EEOC are stretching their mandates to cover algorithmic issues, and if empowered by legislation like the AAA, they could systematically oversee AI practices. A recurring lesson is that regulators must be equipped with **technical expertise** and possibly new investigative powers to audit algorithms (e.g., reviewing source code or datasets), something that traditional legal processes are only beginning to accommodate. The case of Clearview showed regulators willing to use maximum fines and even creative measures (like CNIL's penalty payments), yet it also highlighted difficulties in cross-border enforcement. This suggests that future regulatory strategies will require **greater international cooperation**. For instance, perhaps a treaty

or mutual assistance framework for data protection enforcement, or convergence on principles so that an AI company cannot evade scrutiny by hopping jurisdictions.

Ethically, these case studies drive home that **principles of fairness, transparency, and accountability are not abstract ideals but practical necessities** when AI is integrated into decision-making. Regulation is the tool by which those principles are operationalized. Clearview's misuse of personal data without consent violated privacy and autonomy at a massive scale. Regulatory action asserted the primacy of consent, purpose limitation, and proportionality. SCHUFA's scoring, if left unchecked, risked treating individuals as mere numbers without recourse to the law, is now reasserting the value of human review and agency in consequential decisions. HireVue's scenario warns that efficiency gains from AI mean little if they come at the cost of hidden discrimination or loss of human dignity in the hiring process, prompting regulatory emphasis on **algorithmic fairness and the irreducible role of human judgment**.

Looking ahead, the implications for future regulatory strategies are clear: a **multi-pronged approach** is needed. This includes:

- **Robust legal frameworks** (like GDPR, AI Act, AAA) that cover the different stages and facets of AI systems from data input to algorithm design to decision output and downstream effects.
- **Active enforcement and guidance** by regulators to interpret these laws in concrete cases, as well as to keep regulations up to date with technological advances (e.g., what about new AI like large language models making decisions? The principles gleaned here will need to be applied anew in such contexts).
- **Engagement with stakeholders** regulators should continue to hear from technologists, ethicists, industry, and civil society to understand how AI is used on the ground and where real-world frictions lie. The case studies all involved, in some way, civil society activism

(Privacy International against Clearview, consumer complaints in SCHUFA, EPIC against HireVue) – effective regulation will harness such inputs and not function in isolation.

- **Education and transparency** ensure that individuals know their rights and that organizations know their obligations. For instance, companies might need to invest in training their staff (HR, credit officers, police) on the responsible use of AI tools, guided by regulatory standards. Similarly, public awareness campaigns about rights (like the right to object to automated decisions) can empower those affected to demand compliance.

Lastly, these cases highlight that while laws like GDPR and proposed ones like the AAA are generally adequate in scope (covering key principles needed to govern AI decisions), the proof of adequacy lies in **implementation and enforcement**. A law on paper means little if not enforced, but when enforced, as with GDPR against Clearview or via the CJEU in SCHUFA, it can have a transformative effect on industry behaviour and protection of individuals. Therefore, the ongoing task for regulators and legislators is to ensure these frameworks are not only well-crafted but also well-resourced and assertively applied. The AI landscape is fast-changing; regulatory agility will be essential. But the fundamental goals remain constant: to protect human rights, prevent harm, and promote an AI ecosystem that is **innovative yet trustworthy**. The three case studies of this chapter serve as instructive milestones on that journey, each illustrating pitfalls to avoid and guiding principles to uphold as we strive for adequate governance of AI-powered decision-making in the years ahead.

7. Recommendations and Future Directions

In light of the preceding evaluation of the GDPR and the proposed Algorithmic Accountability Act (AAA) as governance mechanisms for AI-driven decision-making, this chapter provides structured

recommendations for reform. It first proposes **legislative reforms** to strengthen these frameworks (Section 7.1), then offers **comparative policy recommendations** drawing on international models (Section 7.2), and finally outlines **future research directions** to address unresolved issues at the intersection of AI, law, and ethics (Section 7.3).

7.1 Legislative Reform Proposals

7.1.1 Clarifying and Strengthening the GDPR

Clarify the scope of Article 22 GDPR: Article 22, which grants data subjects the right **not** to be subject to certain solely automated decisions, requires clarification to close ambiguity and loopholes. Currently, Article 22's protection is limited by the narrow definition of a "decision based solely on automated processing" and the difficulty of determining what qualifies as a *similarly significant* effect.[207] This narrow scope means that companies can often avoid the Article 22 regime by inserting minimal human involvement (even perfunctory) in automated decision processes. As Helena Vrabec observes, *"all the rights under Article 22 of the GDPR are considerably limited due to the narrow definition of solely automated decisions and the difficulties of setting up an explanatory and contesting process in practice."*[208] To address this, the GDPR should be amended or supplemented with guidance to **broaden the definition** of automated decision-making subject to Article 22. For example, clarifying that decisions with *minimal* human input or rubber-stamping are still "automated" would prevent controllers from evading accountability by nominal human intervention.[209] The ongoing *SCHUFA* credit scoring case before the CJEU underscores this need for clarity: it raised the question of whether generating a credit score (later used by a lender) counts as an automated decision under Article 22.[210] The Advocate General's opinion favoured a broad

[207] Helena U Vrabec, *Data Subject Rights under the GDPR* (OUP 2021) https://doi.org/10.1093/oso/9780198868422.001.0001
[208] Ibid.
[209] Ibid.
[210] Boris Paal, 'Article 22 GDPR: Credit Scoring Before the CJEU' (2023) 4 *Global Privacy Law Review* 127.

interpretation, treating the score itself as a decision, highlighting the importance of **legislative guidance** on such borderline scenarios.[211] Clear statutory or regulatory **criteria for "significant effect"** (e.g. rejection for credit, employment, etc.) and for what constitutes "solely automated" would ensure Article 22 is applied consistently and as intended to truly high-impact AI decisions.

Strengthen transparency and explainability obligations: Hand in hand with clarifying Article 22, the GDPR's transparency requirements (Articles 13–15) should be bolstered to deal with algorithmic opacity. Currently, data subjects have the right to receive "meaningful information about the logic involved" in automated decisions (GDPR Art. 15(1)(h)), but this provision is vague and weakly enforced. The law should explicitly require **ex ante disclosure** when organisations deploy AI that significantly affects individuals, and mandate that intelligible explanations be provided about how the decision was reached. In practice, this means moving toward a true "*right to explanation*", ensuring data subjects can understand the reasons behind an algorithmic decision.[212] This aligns with ethical principles of explicability and transparency advocated in EU guidelines for Trustworthy AI.[213] Concretely, the GDPR or accompanying guidance could require simpler, plain-language explanations of automated outcomes, information on key input factors, and the logic or criteria the algorithm used to arrive at its decision. Such measures would "infuse fairness into automated decision-making processes"[214] by empowering individuals with knowledge and recourse. Importantly, greater transparency must be balanced with trade secrets, so **confidentiality safeguards** (similar to those in the AI Act) can be introduced to protect genuine business secrets while still informing individuals of *their* case particulars.[215]

Reinforce enforcement mechanisms under the GDPR: Even the clearest rights mean little without robust enforcement. The GDPR's enforcement to date in the AI context has been inconsistent across Member States. Empirical analysis shows that while a few DPAs (in France,

[211] Ibid.
[212] Vrabec (n 205)
[213] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (European Commission, 8 April 2019) https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[214] Vrabec (n 205)
[215] Almada (n 83)

Italy, Spain, etc.) have actively pursued cases involving AI or automated decisions, many others have been **"reluctant" or even "inert" enforcers** in this area.[216] To strengthen enforcement, regulators should be **better resourced and coordinated**, and procedural mechanisms improved. One recommendation is to formally designate national Data Protection Authorities (DPAs) as lead enforcers for algorithmic decision-making issues, given their expertise with personal data processing.[217] This could include empowering DPAs (or a specialised unit thereof) to audit AI systems for GDPR compliance (e.g. inspecting algorithmic impact assessments or source code where needed[218]) and to handle complaints about opaque or biased AI decisions. Additionally, an **EU-wide database of AI-related DPA decisions** should be created to promote transparency and consistency in enforcement.[219] Common standards for publishing and sharing these decisions would allow regulators and researchers to identify trends and gaps, and would put companies on notice of what practices are unacceptable.[220] The European Data Protection Board (EDPB) can also issue harmonised **guidelines on AI and GDPR** (e.g. interpreting Article 22 in various scenarios), which DPAs can then enforce uniformly. Finally, higher penalties and more frequent use of GDPR's fine powers for breaches of Article 22 and related provisions would create a deterrent effect, incentivising controllers to take these obligations seriously. In sum, bolstering enforcement through clearer mandates, collaboration, and transparency will give teeth to the GDPR's AI provisions and ensure controllers cannot ignore their responsibilities with impunity.

7.1.2 Refining the EU AI Act Framework

Address vague risk-tiering criteria in the EU AI Act: The EU Artificial Intelligence Act (AI Act) represents a landmark risk-based approach, classifying AI systems into prohibited, high-risk, and lower-risk categories. However, throughout its development, stakeholders noted that the criteria for

[216] Joanna Mazur, Claudio Novelli and Zuzanna Choińska, 'Should Data Protection Authorities Enforce the AI Act? Lessons from EU-wide Enforcement Data' (12 June 2025) SSRN https://ssrn.com/abstract=5290513 or http://dx.doi.org/10.2139/ssrn.5290513
[217] Ibid.
[218] Almada (n 83)
[219] Mazur, Novelli and Choińska (n 214)
[220] Ibid.

classifying systems as "high-risk" were sometimes **vague and under-specified**, leaving room for divergent interpretation.[221] The Act's definition of high-risk AI (initially tied to certain use-case categories in Annex III and the notion of significant harm to health or fundamental rights) should be revisited to **sharpen and clarify the boundaries**. For example, clearer guidelines or delegated acts could define what counts as a "significant adverse effect" on rights or welfare, with concrete examples, to guide providers and regulators. As it stands, some systems may slip through gaps or be misclassified. Civil society groups have pointed out that the Act's list of high-risk use cases and various **exemptions create legal uncertainty and potential loopholes**.[222] Notably, certain AI systems can avoid high-risk scrutiny by claiming to be mere "accessory" tools or by exploiting exceptions (such as for systems aiding human decision-makers rather than replacing them).[223] To prevent abuse, the legislation should tighten these definitions for instance, requiring that any AI system **materially influencing** significant decisions (in credit, employment, policing, etc.) be treated as high-risk unless proven otherwise. The involvement of human users should not automatically downgrade an AI system's risk level if in reality the human oversight is nominal. In short, the EU should refine the risk criteria to be **predictable, context-sensitive, and future-proof**, possibly via an expert rulemaking process that can adapt the Annex III list as new high-risk use cases (like advanced generative AI applications) emerge. This will ensure truly risky AI does not fall outside the Act's stringent safeguards due to definitional gray areas.

Mandate ex ante *independent* audits for high-risk AI systems: Under the current AI Act text, providers of high-risk AI are required to undergo a conformity assessment **before** putting their system on the market (AI Act Art. 19, 43). However, many high-risk AI systems are subject to a form of **self-assessment** (internal checks by the provider) rather than a third-party audit, except in cases involving certain harmonized standards or specific critical uses. To bolster trust and

[221] European Centre for Not-for-Profit Law, Liberties and European Civic Forum, 'Packed with loopholes: why the AI Act fails to protect civic space and the rule of law' (ECNL, 3 April 2024) https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law

[222] Amdouni K, 'Unanswered Concerns on the EU AI Act: A Dead End?' (Global Policy Journal, 3 May 2024) https://www.globalpolicyjournal.com/blog/03/05/2024/unanswered-concerns-eu-ai-act-dead-end; European Centre for Not-for-Profit Law, Liberties and European Civic Forum (n 219)

[223] Amdouni (n 220)

compliance, the Act should introduce **mandatory ex ante audits by independent experts or authorities for all high-risk AI systems**. This would function akin to a required "AI impact assessment" reviewed by an external auditor or regulator prior to deployment. The rationale is that truly high-stakes AI, e.g. an algorithm deciding who is admitted to education, or an AI system used in law enforcement, should not rely solely on the provider's own assurances of safety. Indeed, the final compliance evaluation in the Act heavily depends on providers' stated risk and compliance documentation, including a compulsory Fundamental Rights Impact Assessment (FRIA). Yet **"the precise extent of those responsible for carrying out the FRIA remains somewhat unclear"** and many deployers lack the expertise to evaluate their AI's risks thoroughly. By requiring an independent audit or regulatory approval step, we ensure a qualified review of the AI system's training data, design logic, and mitigation measures *before* harm can occur. This echoes calls by ethicists and scholars for external algorithmic accountability mechanisms, moving beyond pure self-regulation.[224] An *ex ante* audit regime could be implemented by accredited third-party assessors (similar to notified bodies in product safety law) or by an empowered supervisory authority (e.g. national AI regulators). Such audits should evaluate not only technical criteria (accuracy, robustness) but also check for compliance with **fundamental rights and non-discrimination requirements**. By making external audits mandatory, the Act would add a robust layer of oversight, ensuring that providers' internal risk assessments (which may underestimate problems) are verified. This measure complements the above call for clearer risk definitions: once it's clear which systems are high-risk, those systems would uniformly undergo a rigorous audit *before* they impact users, rather than relying on voluntary compliance. Overall, **embedding independent scrutiny in the development lifecycle** of high-risk AI will greatly enhance accountability and trust in AI systems placed on the EU market.

Impose limits on voluntary Codes of Practice and strengthen binding rules: The EU has considered voluntary Codes of Conduct (or practice) for AI, especially for AI systems not classified

[224] Mökander, Juneja, Watson and others (n 98).

as high-risk. While collaborative industry codes can be useful for spreading best practices, they **must not become a substitute for binding obligations** in areas that truly matter for fundamental rights. Critics warn that the AI Act "leaves too much to… voluntary codes of conduct," which could **"undermine the safeguards"** established by law.[225] Indeed, voluntary *general-purpose AI* Codes (the Commission has promoted one for foundation models) are only effective if widely adopted and strictly adhered to – conditions far from guaranteed. The final law should therefore impose **clear limits on the role of voluntary codes**: for instance, codes should *supplement* but not replace regulatory requirements, and there should be consequences for non-compliance by those who sign up. One recommendation is to integrate a mechanism by which code commitments can be made **enforceable**. If an AI provider claims adherence to a Code of Conduct (e.g. on fair AI or on generative AI safety), that commitment could be made legally binding under the Act's enforcement (perhaps via Article 69(2) which allows the Commission to recognize such codes). Additionally, the Act's reliance on delegated acts and future guidelines to fill gaps should be balanced with stronger baseline rules in the text. The rushed negotiations of the Act left **"significant gaps and legal uncertainty"** on certain protections; voluntary codes or future guidance are expected to patch these, but that approach is fragile. The **risk-tier system itself should be preserved as binding**, meaning if new risky practices emerge (like emotion recognition in sensitive contexts), regulators should not rely solely on a voluntary code to manage them, but rather update the law or use enforcement powers. Finally, to avoid purely self-regulatory approaches, the Act's provisions for **stakeholder participation** and civil society input should be strengthened (e.g. requiring that any code of conduct be developed in consultation with fundamental rights organisations and be subject to approval by the European AI Board). By ensuring that voluntary measures do not become "smoke and mirrors" but are anchored in accountability, the EU can prevent a scenario where industry codes dilute or delay real compliance. The bottom line is that **binding obligations must remain primary**, especially for anything affecting rights, with voluntary codes serving only as a complementary tool to raise standards above the legal minimum, not as an escape hatch from regulation.

[225] European Centre for Not-for-Profit Law, Liberties and European Civic Forum (n 219)

7.1.3 Enhancing the U.S. Algorithmic Accountability Act (AAA)

Expand the AAA's scope to include SMEs and high-risk sectors: The U.S. Algorithmic Accountability Act (AAA), as proposed in 2022 and reintroduced in 2025, is a promising federal effort to mandate algorithmic impact assessments, but its applicability criteria are currently too limited. The bill defines "covered entities" mainly by size (e.g. businesses over a certain revenue or data processing threshold).[226] This leaves out many small- and medium-sized enterprises (SMEs) that increasingly develop or deploy AI systems with significant impacts (for instance, a small EdTech startup using AI to evaluate students, or a hiring tool used by a mid-size firm). Harm caused by automated decisions is not limited to Big Tech alone. **The AAA should broaden its scope** so that high-impact AI uses are regulated regardless of the size of the deploying entity. One approach is to include **all operators in designated high-risk sectors or use-cases**, even if they don't meet the current revenue or data thresholds. For example, any company (regardless of size) using AI for employment screening, creditworthiness assessment, insurance underwriting, healthcare triage, or other sensitive decisions could be brought under the Act's requirements. This ensures that critical applications affecting individuals' livelihoods or rights are subject to algorithmic accountability measures across the board. Expanding the scope aligns with the risk-based approach: **if an AI system is high-risk, the law applies**, period. The Act could maintain some proportionality by tailoring requirements (e.g. simpler impact assessments for micro-enterprises), but it should not categorically exempt all SMEs. Without this change, there is a danger of **regulatory gaps** where exactly those innovative smaller companies who may lack internal ethical oversight deploy unchecked algorithms. Notably, lessons from other jurisdictions (like the EU AI Act) show the importance of covering *all* providers of high-risk AI, with some support for small players rather than exclusion. By extending its scope, the AAA would close loopholes and prevent a two-tier system where big firms must assess algorithmic harms but startups get a free pass even when operating in high-risk domains.

[226] AAA 2025 (n 109)

Improve stakeholder engagement requirements for algorithmic governance: The AAA, as currently drafted, breaks new ground in requiring organisations to consider input from stakeholders during algorithmic impact assessments. It instructs "covered entities" to engage with **relevant internal and external stakeholders**, including employees, ethics teams, and representatives of impacted groups, and even to document how stakeholder feedback was used or why it was not incorporated.[227] This is a valuable mechanism for democratizing AI oversight. However, to ensure this is not a mere box-ticking exercise, the AAA's stakeholder engagement provisions should be **made more rigorous and binding**. First, the Act should clarify *which* stakeholders must be consulted for various contexts, for instance, requiring input from civil rights or consumer advocacy groups when an AI system affects those constituencies. It should also set a **minimum standard for the consultation process**: simply hosting a brief meeting may not suffice. Instead, companies could be required to solicit public comments or hold workshops with affected communities and **meaningfully integrate the feedback** into design changes. The Act could mandate that the resulting impact assessment report include a section detailing stakeholders' concerns and the company's responses, ensuring accountability. As one analysis noted, the AAA would make companies "explain their rationale for setting aside stakeholder input"[228] This is a good start, but regulators should have the power to review whether dismissing certain stakeholder recommendations was justified or if the company ignored red flags. Additionally, the law might encourage creation of **advisory committees** or external ethics boards for high-risk AI projects, embedding multi-stakeholder governance into the AI lifecycle. Empirical research supports such engagement: involving diverse perspectives leads to more robust identification of risks, including those the developers might overlook.[229] By tightening and emphasizing the stakeholder engagement requirements, the AAA would not only check the corporate discretion in AI development but also empower affected communities to have a say in how AI systems are designed and deployed.[230] This

[227] Ibid.
[228] Maroussia Lévesque, 'Smoke and Mirrors? AI Regulation and Corporate Power' (2024) *University of Illinois Journal of Law, Technology & Policy* (2025) ISS 1 https://ssrn.com/abstract=4736131
[229] Amdouni (n 220)
[230] Ibid.

participatory approach is in line with emerging global norms for AI ethics and will bolster the legitimacy of algorithmic decision-making by ensuring it accounts for a plurality of values and concerns.

Ensure binding obligations and enforcement under the AAA: A key concern with AI governance in the U.S. is the current lack of binding rules much relies on voluntary guidelines or sectoral enforcement. The AAA is meant to change that by giving the Federal Trade Commission (FTC) authority to enforce new rules on algorithmic accountability. To fulfill this goal, the final Act must guarantee that its obligations are truly **binding and enforceable**, not weakened by loopholes or discretionary enforcement. This entails several components. First, the FTC's rulemaking under the Act should be time-bound and comprehensive, yielding clear regulations that companies *must* follow (e.g. specifying the format and methodology of algorithmic impact assessments, the necessity of bias testing, etc.). Second, the Act should provide the FTC with adequate enforcement powers – including the ability to levy substantial fines for non-compliance, to require cessation of problematic algorithms, and to impose remedies (such as algorithmic audits or notice to consumers). Currently, the FTC can already investigate unfair or deceptive practices; the AAA would expand this to algorithmic harms. It is crucial that this expansion comes with **sufficient funding and expertise** for the FTC to carry out audits of AI systems and not rely solely on companies' paperwork. Third, **remove or minimize any safe harbors** that could render the obligations toothless. For example, if the Act allowed companies to comply via **voluntary codes of conduct** (similar to the EU approach) or gave too much leeway in defining "reasonable" measures, it might result in minimal changes. Instead, requirements like conducting impact assessments, addressing identified risks, and periodically reporting on AI practices should be unequivocal duties, not recommendations. Comparative insight from Canada's privacy law reform is instructive here: the Office of the Privacy Commissioner (OPC) has called for stronger powers, including the ability to make binding orders and impose penalties under PIPEDA, moving away from soft guidance to hard enforcement. The AAA likewise should not remain a mere *proposal* or an aspirational

framework. Congress should **formally enact it into law** and ideally complement it with a private right of action or state enforcement to back up FTC action. In sum, to avoid the pitfalls of previous self-regulatory approaches, the AAA must emerge as a statute with clear, binding rules for algorithmic governance and a strong enforcement scheme. If implemented as such, it would mark a significant shift in the U.S., aligning it more closely with the EU's rights-based approach (while still allowing flexibility through an outcomes-focused, assessment-based process). Ensuring these obligations are binding will increase corporate compliance: companies will know that failure to assess and mitigate AI risks or to honestly engage stakeholders **will have legal consequences**, not just reputational ones.

7.2 Comparative Policy Recommendations

In crafting the above reforms, it is instructive to **draw lessons from international frameworks and best practices** in data protection and AI governance. This section compares approaches from the OECD, APEC, and Canada's PIPEDA, recommending how their strengths can be integrated to improve GDPR, the EU AI Act, and the AAA.

7.2.1 Lessons from OECD Guidelines and Global AI Principles

One of the oldest and most influential sets of data governance principles comes from the **OECD Privacy Guidelines** (originally of 1980, updated in 2013). The OECD Guidelines introduced foundational concepts such as purpose limitation, data quality, security safeguards, and the **accountability principle** which have shaped laws like the GDPR.[231] A key comparative insight is the importance of **accountability**: under the OECD approach, organizations that transfer or handle personal data are expected to ensure protection regardless of jurisdiction, rather than relying solely on governments to pre-clear data flows.[232] This has parallels in AI governance: organisations deploying AI should be accountable for compliance and ethical outcomes, even in the absence of

[231] Mark Phillips, 'International Data-Sharing Norms: From the OECD to the General Data Protection Regulation (GDPR)' (2018) 137 *Human Genetics* 575 https://doi.org/10.1007/s00439-018-1919-7
[232] Ibid.

detailed prescriptive rules. Incorporating this, regulators in the EU and US should emphasize organizational accountability for AI for instance, by requiring internal compliance programs, ethics committees, and continuous monitoring of AI impacts (a concept akin to the accountability-based model in international privacy regimes)[233].

Another lesson from the OECD is the value of **global consensus principles**. The **OECD AI Principles (2019)** endorsed by 42 countries articulate high-level norms for trustworthy AI: inclusive growth, human-centered values, transparency, robustness, and accountability. These mirror the EU's own ethics guidelines[234] and converge with frameworks like Floridi et al.'s "five principles" (beneficence, non-maleficence, autonomy, justice, explicability).[235] Such global principles indicate a **convergence on core values** for AI. It is recommended that lawmakers explicitly reference or incorporate these principles in explanatory memoranda and regulatory guidance, to ensure a common vocabulary and goal. For example, the EU could ensure that its AI Act implementation guidance aligns with OECD AI Principles on transparency and risk management. The US, through the AAA or subsequent rules, can similarly cite these principles to guide interpretation (the AAA's focus on **mitigating "social, ethical, and legal" risks of AI** is very much in the spirit of OECD's human-rights-based approach).[236] Aligning with international principles not only lends legitimacy but also facilitates interoperability: AI developers operating globally can aim to meet one coherent set of expectations.

Finally, the OECD experience underlines the necessity of **multilateral cooperation** for governing digital technologies. Just as OECD privacy norms sought to harmonize protection to enable data flows,[237] today's regulators should collaborate on AI standards. The EU's participation in the *Global Partnership on AI (GPAI)* and the US-EU Trade and Technology Council's AI initiatives are positive steps. This chapter recommends further support for developing **international standards**

[233] Ibid.
[234] HLEG on AI, Ethics Guidelines for Trustworthy AI (n 211)
[235] Floridi and Cowls (n 93)
[236] Mökander, Juneja, Watson and others (n 98)
[237] Phillips (n 229)

for AI audits, transparency reporting, and risk assessments under bodies like ISO/IEC or IEEE, inspired by OECD's convening role. In summary, by learning from OECD frameworks, policymakers can reinforce **accountability and common principles** in AI regulation, helping to raise the floor of AI governance worldwide.

7.2.2 Insights from APEC CBPRs and Cross-Border Accountability

In the Asia-Pacific, a noteworthy model is the **APEC Cross-Border Privacy Rules (CBPR) system**, which offers a **voluntary but formalised** framework for data protection across jurisdictions. Under APEC's Privacy Framework (2005), the emphasis is on the *accountability principle* organisations certify their compliance with baseline principles and are held accountable via recognised Trustmark agencies.[238] The **advantage** of this approach is flexibility and interoperability: it allows data to flow while relying on businesses to implement privacy programs that meet agreed standards. However, a critical observation is that **APEC CBPR is voluntary and lacks independent legal effect**, limiting its uptake and enforcement power.[239] Only a handful of APEC economies and companies have participated meaningfully, showing the difficulty of reliance on voluntary certification alone.[240]

The lesson for AI governance is twofold. First, the **accountability-based model**, where an entity exporting an AI system or algorithmic service ensures it will be handled according to certain safeguards by importers, could inspire frameworks for cross-border AI deployments. For example, if AI systems trained in one country are used in another, an APEC-like mutual recognition of **algorithmic accountability certifications** could facilitate trust. Regulators could encourage firms to adopt **AI management systems** that are audited and certified for fairness and security, easing global adoption. This might be particularly useful for emerging international AI standards or an idea of a global AI compliance seal (akin to CBPR's trustmark).

---

[238] Ibid.
[239] Ibid.
[240] Ibid.

Second, the **limitations of purely voluntary schemes** counsel that any such approach must be backed by the possibility of enforcement. In the APEC CBPR, while joining is voluntary, once a company commits, it can face consequences from domestic regulators if it breaches the certified rules.[241] Similarly, for AI, voluntary **Codes of Conduct** (as discussed for the EU AI Act) should be tied to regulatory oversight – e.g. authorities monitoring compliance or integrating code adherence into law (making it enforceable). The U.S. could consider leveraging the AAA to recognize industry AI codes, but only with FTC oversight to ensure they are not mere window dressing. The comparative point is that **voluntarism works only with accountability and external scrutiny**. APEC's experience showed limited initial engagement, but also a "sustained commitment" to updating the system and involving enforcement bodies.[242]

Therefore, policymakers are advised to adopt the **strengths** of the APEC approach (accountability, interoperability) for instance, encouraging *organizational accountability programs for AI* that transcend borders while mitigating its weaknesses by making such programs **opt-in but binding once opted-in**. Furthermore, cross-regional dialogues (EU-APEC, etc.) can explore mappings between CBPR and GDPR (the two share common principles historically[243]) and perhaps in the future, between CBPR and the EU AI Act's provisions. The eventual emergence of a **Global CBPR Forum** and discussions at the OECD on AI indicate momentum for multilateral solutions. In conclusion, **APEC's model teaches the value of accountability and multistakeholder design**, but also highlights that to truly protect rights, voluntary compliance frameworks must be coupled with strong oversight and not used as a substitute for mandatory protections.[244] Blending these insights can help create an AI governance ecosystem that is both nimble and trustworthy across borders.

[241] Ibid.
[242] Ibid.
[243] Ibid.
[244] Ibid.

7.2.3 Adopting Best Practices from Canada's PIPEDA

Canada's private-sector privacy law, the **Personal Information Protection and Electronic Documents Act (PIPEDA)**, offers a *principles-based, flexible approach* that can inform AI governance. PIPEDA is built on ten Fair Information Principles (including accountability, openness, and individual access), rather than highly detailed rules. This flexibility has allowed it to adapt to new challenges without constant legislative change, a useful trait when facing rapidly evolving AI technology.[245] For instance, although PIPEDA does not explicitly name "automated decision-making" in its original text, its broad principles have been interpreted to cover algorithmic transparency to some extent. The Office of the Privacy Commissioner of Canada (OPC) has clarified that organisations must be transparent about their use of automated decision systems and, upon request, explain to individuals how a decision about them was made (a right that was to be codified in the newer Consumer Privacy Protection Act proposal).[246] This indicates a **proactive interpretation of existing principles to new technology,** something from which EU DPAs and U.S. regulators can learn. Rather than waiting for explicit new laws, they can use existing legal principles (transparency, fairness, etc.) to set expectations for AI. For example, European regulators might use GDPR's fairness and transparency requirements (Arts. 5(1)(a), 12) to similarly demand explanation of algorithmic decisions even before Article 22 is clarified, much as the OPC has used PIPEDA's principles in guidance.

Another best practice from Canada is the OPC's **proposals for reforming PIPEDA to address AI**, which were put forward in 2020. These included: an explicit definition of automated decision-making in law; a right for individuals to receive an explanation of automated decisions affecting them; requirements for **algorithmic auditing or impact assessments**; and perhaps most critically, stronger enforcement powers for the OPC (such as the ability to issue binding orders and monetary penalties). The rationale was to modernise a principles-based law with specific provisions

[245] Office of the Privacy Commissioner of Canada, A Regulatory Framework for AI: Recommendations for PIPEDA Reform(OPC, November 2020)
[246] Office of the Privacy Commissioner of Canada, Policy Proposals for PIPEDA Reform: Automated Decision-Making and Accountability Module (OPC, November 2020)

for AI risks and to give it "teeth" comparable to the GDPR. The recommendation here is to **integrate similar elements in other regimes**: for the GDPR and EU AI Act, ensure individuals have a clear right to explanation and contestation (as discussed in 7.1.1), and for the AAA, ensure regulators have order-making power and penalties (as discussed in 7.1.3). PIPEDA's focus on **stakeholder consultation** during its reform (the OPC engaged public and experts on AI issues[247]) is also a model to emulate: the EU could involve multidisciplinary experts in developing AI Act guidance, and the U.S. could convene expert panels under the FTC to refine AAA rules.

Moreover, Canada's balanced approach, protecting privacy while acknowledging "legitimate business interests" and innovation,[248] provides a middle ground perspective. It suggests that AI regulation can protect individuals *and* permit responsible innovation by being principles-based and collaborative. For instance, PIPEDA allows for **sectoral codes of practice** (which can be formally recognised) as a way to implement its principles in context. The EU might similarly encourage **sector-specific AI codes** aligned with the AI Act, and the U.S. could allow industry standards to fulfil AAA obligations where appropriate, provided they meet baseline criteria. The crucial caveat from Canada's experience is enforcement: historically, PIPEDA relied on moral suasion, but now there is consensus that stronger enforcement is needed (Bill C-27, currently under debate, reflects this by proposing administrative fines for privacy violations, including automated decision transparency failures). Thus, any incorporation of Canada's flexible, principle-driven approach should be coupled with **firm enforcement mechanisms** to avoid the pitfall of laws that look good on paper but are ignored in practice.

In summary, **comparative analysis** suggests blending the **European emphasis on fundamental rights and strict enforcement** with the **flexibility and accountability focus of frameworks like OECD, APEC, and PIPEDA**. This combined approach would help create AI governance regimes

[247] Office of the Privacy Commissioner of Canada, *OPC Leadership on AI and Privacy* (OPC, 6 May 2025) https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/leadership-ai/

[248] Office of the Privacy Commissioner of Canada, *Artificial Intelligence Consultation – Final Submissions* (OPC, November 2020) https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/pol-ai_202011/

that are robust yet adaptable, globally interoperable, and continuously guided by ethical principles and empirical learnings.

7.3 Future Research Directions

While regulatory reforms will address many gaps, several **unresolved issues at the frontiers of AI, law, and ethics** require further research. This section highlights key areas where interdisciplinary inquiry is needed to support evidence-based policy in the future.

7.3.1 AI Explainability and Meaningful Transparency

One pressing research question is how to achieve **AI explainability** that is legally and practically meaningful. Current laws (GDPR, upcoming AI Act) gesture toward a right to information or explanation, but what constitutes an adequate explanation for a complex machine learning model remains debated. Technically, methods like LIME or SHAP can provide post-hoc explanations of black-box models, but their outputs (e.g. feature weight plots) may not be intelligible to lay data subjects. Meanwhile, simply disclosing the source code or algorithmic parameters would overwhelm users and reveal little about *why* a decision was made.[249] There is a need for research into **socio-technical approaches** to explanation: for example, combining user experience research (to find out how people best understand algorithmic outcomes) with computer science (to generate narratives or visualisations of an AI's reasoning). Studies could explore how to convey the logic of different AI models (rules-based vs. neural nets) in plain language, and how much information is optimal to avoid information overload versus opacity.[250] Additionally, scholars are examining the concept of *"explanation accuracy"*: ensuring the explanation is a faithful description of the actual model decision process, not a misleading simplification. Regulators will need guidance on setting standards for explanations under laws like Article 22 GDPR or Article 52 of the AI Act (which requires some transparency for high-risk AI). Therefore, interdisciplinary research involving

[249] Almada (n 83)
[250] Vrabec (n 205)

computer scientists, legal theorists, cognitive psychologists, and HCI experts should focus on developing **techniques for "intelligible AI"** and testing them in user studies. The goal is to answer questions like: What do individuals *want* to know about an algorithm's decision? What formats of explanation increase trust or enable effective contestation? How can we ensure explanations do not expose trade secrets unnecessarily, yet still uphold the individual's right to understand? These inquiries will directly inform future regulatory guidance and standards (possibly even technical standards under the AI Act's Annexe IV, which could detail how transparency info should be provided). In sum, making AI transparent in practice is as much a human factors challenge as a technical one, and continued research is vital to bridge the gap between legal transparency rights and black-box AI technology.

7.3.2 Algorithmic Fairness Metrics and Nondiscrimination

Another area for future research is developing and reconciling **algorithmic fairness metrics** with legal and ethical concepts of fairness. Over the past decade, computer scientists have proposed dozens of quantitative definitions of "fairness" (e.g. demographic parity, equalized odds, predictive value parity), but it is mathematically impossible to satisfy all definitions simultaneously in a single model except in trivial cases.[251] This has been dubbed the *fairness impossibility theorem*.[252] For regulators and practitioners, this raises the question: which definition of fairness should an AI strive for, especially given different legal norms across jurisdictions? As noted, *"many of the metrics proposed for algorithmic fairness do not tackle the same problems required by EU law"*.[253]European non-discrimination law, for example, allows indirect discrimination claims even without intentional bias, and values context-specific justification over purely statistical parity. U.S. law, by contrast, often focuses on disparate impact analysis involving specific protected classes. Research should examine how existing fairness metrics align or conflict with these legal frameworks. It may be that entirely new metrics or multi-metric approaches are needed to capture

[251] Almada (n 83)
[252] Ibid.
[253] Ibid.

legally relevant unfairness (for instance, a metric that considers counterfactual fairness: would the decision be different if the person belonged to a different demographic group?). Furthermore, technical research could explore **algorithmic tools for bias mitigation** e.g. pre-processing data to remove proxies for protected attributes, or post-processing outputs to achieve a chosen fairness criterion and evaluate their efficacy and any unintended consequences.

An important interdisciplinary direction is to engage with **philosophical and sociological perspectives on fairness**: fairness is not purely numerical, and cultural values influence what outcomes are seen as just. Hence, legal scholars and ethicists should work with data scientists to frame metrics that reflect societal values (such as fairness as justice, which might align with the principle of *equality* highlighted by Vrabec as a fundamental value[254]). There is also a need for research on **bias in AI across different domains**: fairness in criminal justice risk assessments might require different considerations (e.g. due process, avoiding reinforcing historical biases in policing) compared to fairness in, say, credit lending or recruitment. Sector-specific studies can inform regulators where bright-line rules make sense versus where flexibility is needed. As the EU AI Act moves toward implementation, it mandates that high-risk AI systems ensure **non-discrimination** (Article 10 on data and Article 13 on accuracy). Yet, without concrete methods to test and certify fairness, these obligations may remain abstract. Ongoing research, potentially guided by initiatives like the EU's Horizon Europe programs or NIST's AI Risk Management Framework in the U.S., should aim to produce **standardised fairness testing methodologies**. Ideally, future interdisciplinary work will yield *practical guidelines* for developers (e.g. how to conduct a bias audit) and inform future revisions of laws to include explicit fairness requirements grounded in empirical evidence. Ultimately, the challenge of algorithmic fairness is as much a social debate as a technical one research must therefore remain attuned to public values and the lived experiences of communities impacted by AI-driven discrimination.

[254] Vrabec (n 205)

7.3.3 Human Oversight and Effective Human–AI Teaming

The principle of **human oversight** is a common requirement in AI ethics and now law (it is one of the seven key requirements in the EU's ethics guidelines and features in the AI Act, which requires human oversight for high-risk AI, e.g. Article 14). However, what *"meaningful human oversight"* entails is still an open question. Research is needed on designing oversight mechanisms that truly empower humans to correct or override AI, rather than simply rubber-stamping automated outcomes; thus, the AI Act's language of requiring oversight when technically feasible introduces ambiguity and potential loopholes. Empirical study is required to determine when and how human intervention is feasible and effective. For instance, in high-frequency trading algorithms, human real-time oversight might be technically infeasible due to speed, suggesting oversight must come in the form of controls and circuit breakers rather than per decision review. In other contexts, like automated hiring tools, human recruiters could, in theory, vet each algorithmic recommendation, but if they almost always defer to the AI, the oversight is illusory. This phenomenon of automation bias (humans over-relying on AI suggestions) is well-documented and deserves further investigation under various conditions. Researchers should experiment with interface designs and decision protocols to mitigate automation bias, e.g. by concealing the AI's prediction confidence or by forcing human reviewers to make an initial judgment before seeing the AI's output.

Additionally, there is a need for **research on training and guidelines for human overseers**. What kind of training helps a human operator detect when an AI might be in error or unfair? Are there psychological or organisational factors (like workload, trust in technology, corporate culture) that influence the effectiveness of oversight? Interdisciplinary studies (combining cognitive psychology, organisational science, and human-computer interaction) can yield insights into how to keep humans in the loop in a *meaningful* way. The goal should be to develop models of *human-AI teaming* where each complements the other's strengths, AI for consistency and speed, human for contextual judgment and ethical considerations. In safety-critical fields (like healthcare or aviation),

decades of research on human-in-the-loop systems can be applied and furthered for modern AI contexts.

From a legal perspective, scholars might explore comparative oversight models: for example, **human-in-command** approaches (where a human has final decision power, as in GDPR's Article 22(3) "right to human review") versus **human-on-the-loop** (where humans monitor decisions and can intervene during operation) versus **human-in-the-loop** (where each decision involves human confirmation). Each model has implications for liability and responsibility. Ongoing discussions about the **AI Liability Directive in the EU** also raise the question of how to assign responsibility when AI and humans jointly make a decision. Research in this area might directly inform not only best practices but also future amendments to laws to clarify accountability in human-AI teams.

In summary, ensuring that human oversight is not just nominal requires a deeper understanding of the human factors. Future research should strive to produce **evidence-based guidelines** for effective oversight, which legislators and companies can adopt. This might include recommended limits on full automation in certain sensitive scenarios (banning "human-out-of-the-loop" in, say, life-altering decisions), and tools to enhance human control (such as decision support dashboards that highlight anomalies in the AI's reasoning for human reviewers). By addressing these questions, we can support the development of oversight regimes that genuinely uphold the principle of human autonomy and control in AI deployments, rather than providing a false sense of security.

7.3.4 Interdisciplinary Collaboration Linking Law, Ethics, and Design

Finally, a broad future research direction is the fostering of **interdisciplinary studies that integrate legal analysis, ethical inquiry, and AI system design**. AI governance is not solely a legal problem or a technical one; it sits at the intersection of normative values and engineering realities. A deeper collaboration between fields is needed to operationalise ethical principles into technical

requirements (the very challenge the EU's Trustworthy AI framework grapples with)[255]. This means, for example, that legal scholars and ethicists should work with computer scientists in the early stages of AI development to bake compliance and ethics *into system architectures* ("ethics/privacy by design"). Conversely, technologists need guidance from the humanities and social sciences to understand the societal context and potential impacts of their systems. **Research hubs and funding programs** that encourage such cross-pollination will be crucial. We see early efforts: some universities now have joint AI policy centres, and initiatives like AI4Good or the global **IEEE Ethically Aligned Design** involve diverse experts. But more systematic integration is required, especially in curriculum (training a new generation of "AI lawyers" and "ethical engineers") and in project funding (e.g. grants that require a legal-ethical work package alongside technical development).

Key topics for this interdisciplinary approach include: **Ethical design patterns**: Identifying design choices that embed values like fairness or privacy (for instance, incorporating explanation facilities and audit logs into AI systems by default). **Legal oversight of AI design**: Studying how regulatory agencies might audit algorithms or data sets and what tools they would need – this is both a legal question (what authority and procedures) and a technical one (what metrics and access). **Impact of AI on legal norms**: As AI becomes more prevalent, it may challenge existing legal notions (such as intent, causation, or even personhood in the case of autonomous systems). Scholars should explore whether concepts like **algorithmic transparency, data ownership, or AI liability** necessitate new legal doctrines or extensions of current ones. For example, should there be a fiduciary duty for companies deploying AI, akin to a doctor's duty, given the potential for harm? These normative questions require philosophical and legal analysis informed by an understanding of AI's capabilities and limits.

Furthermore, global governance issues such as the prospect of an **international AI regulatory regime** or treaty will need inputs from international law experts, diplomats, and technologists to

[255] HLEG on AI, Ethics Guidelines for Trustworthy AI (n 211)

navigate. The role of standards bodies (ISO, IEEE) in quasi-regulating AI through technical standards is another fruitful area of study bridging law (standards referenced in law) and engineering (standards development).[256]

In essence, the future of AI governance research should break silos: law and policy researchers must stay updated on technical advances (e.g. the implications of new AI like GPT-4 on privacy and IP), and AI developers must be cognizant of the legal and ethical environment their systems operate in. This interdisciplinary synergy will ensure that as AI evolves (with new paradigms like federated learning, neuromorphic computing, etc.), our legal and ethical frameworks evolve in tandem, rather than lagging behind by years. Only through continuous collaborative inquiry can we hope to craft governance mechanisms that are resilient to the fast pace of AI innovation while firmly rooted in human rights and societal values.

8. Conclusion

This thesis set out to evaluate how adequately the EU's General Data Protection Regulation (GDPR) and the United States' Algorithmic Accountability Act (AAA) govern AI-powered decision-making. The research question asked whether these frameworks one a comprehensive data protection regime and the other a targeted algorithmic accountability proposal – sufficiently address the legal and ethical challenges posed by AI-driven decisions. Throughout the chapters, the thesis traced the historical development of each framework, dissected their key principles (from lawful bases and data minimisation to transparency and risk-based oversight), compared their scope and enforcement mechanisms, and examined critical gaps and case studies. By addressing these facets in turn, the thesis has answered the question by demonstrating that while both GDPR and the AAA contribute important pieces to AI governance, neither alone is fully adequate; effective regulation of

[256] Mia Hoffmann, 'AI Safety under the EU AI Code of Practice — A New Global Standard?' (Center for Security and Emerging Technology, 30 July 2025) https://cset.georgetown.edu/article/eu-ai-code-safety/

AI decision-making requires both strengthening existing provisions and pursuing new, harmonised approaches grounded in fundamental rights.

The analysis yielded several significant findings. First, the GDPR provides a robust foundation of **data protection principles** including legality, fairness, transparency, purpose limitation, and data minimisation that apply to many AI use cases involving personal data.[257] Its enforcement mechanism (backed by stiff fines) has given teeth to ethical principles like accountability and privacy, even spurring debates about innovation: some critics argue GDPR's strict requirements may slow AI innovation, while others counter that they channel innovation toward privacy-preserving and socially beneficial directions.[258] Notably, GDPR's Article 22, which gives individuals the right not to be subject to solely automated decisions with significant effects, embodies a **rights-based safeguard** directly relevant to AI. However, this protection has proven narrow in practice it applies only to fully automated decisions and carries broad exceptions, making it a somewhat weak shield. The thesis highlighted that ambiguity through the SCHUFA credit scoring case, where it was initially unclear if a score generated by an algorithm (and used by a human creditor) fell under Article 22. In a landmark clarification, the Court of Justice of the EU ruled that even ostensibly "human-in-the-loop" decisions can count as automated decisions if the human's role is essentially to rubber-stamp an algorithmic score thereby bringing such practices within Article 22's ambit.[259] This finding underscores that GDPR, while principled, is still evolving through interpretation and enforcement to meet AI's challenges.

In contrast, the Algorithmic Accountability Act represents a **risk-based regulatory approach** aimed specifically at AI and automated systems. As proposed in the U.S., the AAA would compel companies above certain size thresholds to perform algorithmic impact assessments for their high-risk automated decision systems both before deployment and periodically thereafter and to

[257] Maciej Kuziemski and Przemysław Pałka, *AI Governance Post-GDPR: Lessons Learned and the Road Ahead* (European University Institute, 2019) School of Transnational Governance Policy Brief Series, Issue 2019/07 https://doi.org/10.2870/470055
[258] Ibid.
[259] Almada (n 83)

mitigate any identified harms. This shifts the focus from individual rights to organizational accountability, requiring documentation, transparency reports, and oversight by the Federal Trade Commission. The comparative analysis revealed a sharp divergence in approach: whereas the GDPR embeds AI governance into a broad data protection and fundamental rights framework, the U.S. has thus far addressed AI in a piecemeal way. Without a comprehensive federal privacy law or an enacted AAA, the current U.S. landscape is a patchwork of sectoral rules and state laws, leaving significant gaps. The case studies vividly illustrated these gaps. For example, Clearview AI's facial recognition system which scraped billions of images without consent was aggressively prosecuted under the GDPR by European regulators (resulting in fines and orders to delete data), yet in the U.S. it operated largely unfettered except for isolated lawsuits and state-level bans. This disparity shows the GDPR's extraterritorial strength but also exposes enforcement limits: despite EU fines, Clearview's practices revealed how novel AI services can exploit grey areas (e.g. publicly available data) and how **transatlantic regulatory incoherence** can undermine effective governance. Similarly, the thesis discussed **algorithmic hiring** tools like HireVue: in Europe such tools must comply with GDPR's transparency, fairness and bias mitigation duties, while in the U.S. their use has been largely self-regulated, leading to civil society interventions. In HireVue's case, an NGO complaint to the FTC led the company to halt the controversial facial analysis component of its assessments, implicitly acknowledging that without clear legal standards, public pressure and general consumer protection law filled the void. These case studies underscored the critical gaps when regulatory frameworks are either absent or not rigorously enforced from the weakness of Article 22's initial formulation to the incomplete patchwork of U.S. AI oversight and thereby informed this thesis's recommendations.

Reflecting on the current and future landscape of AI governance, it is evident that we are at an inflection point. **Regulators are moving toward more comprehensive regimes**: the EU is on the cusp of enacting the Artificial Intelligence Act, which will complement the GDPR by imposing tailored requirements on high-risk AI systems (e.g. in credit, employment, policing) and by

introducing obligations like transparency and robustness that extend beyond personal data protection. This development, alongside judicial clarifications (as in SCHUFA) and updated guidance, suggests the EU is knitting a multi-layered framework that combines rights-based protection with product safety-style regulation. In the U.S., the future of AI governance will depend on whether proposals like the AAA become law and how they interplay with existing civil rights and consumer protection laws. The transatlantic comparison in this thesis highlights that *neither* jurisdiction has a perfect solution yet: the GDPR's strengths in empowering individuals and controlling data use need to be bolstered by more explicit AI-focused rules (for example, clearer rights to explanation and recourse in automated decisions), while the U.S. needs to move beyond sector-specific and voluntary guidelines to a more unified, enforceable baseline for AI accountability. Another key insight is the importance of **enforcement and oversight capacity**. Even the best-written law means little if regulators lack the mandate or resources to monitor AI systems. As noted, enforcement of GDPR's AI-related provisions (like Article 22 or mandatory data protection impact assessments) has been patchy, varying widely by country, and U.S. regulators like the FTC are only beginning to flex their muscles on AI issues. Going forward, investing in regulatory capacity from upskilling Data Protection Authorities to possibly creating new oversight bodies (as some have proposed in the U.S.) will be crucial to ensure that the lofty principles in both GDPR and the AAA translate into real accountability on the ground.

In conclusion, the thesis finds that **a harmonised and rights-based framework for AI governance is urgently needed** to address the shortcomings identified. The GDPR and the AAA each embody important aspects of such a framework: the GDPR roots AI regulation in fundamental rights and individual empowerment, whereas the AAA emphasises systematic risk assessment and organizational responsibility. The most effective governance regime will merge these strengths. This means anchoring AI oversight in core principles of **human dignity, non-discrimination, privacy, and transparency** values that the GDPR enshrines while also incorporating the **risk-based, preventative measures** seen in the AAA's approach. The research strongly supports

the view that AI regulation should not be left to a fragmented patchwork or to the goodwill of private companies. Instead, it calls for international coordination and the development of interoperable legal standards. AI-driven decisions often affect individuals across borders, raising global concerns; accordingly, lawmakers and regulators worldwide should strive for greater alignment in their rules and objectives. Encouragingly, there is growing recognition of this need. For instance, Europe's High-Level Expert Group on AI has urged a *global framework for trustworthy AI* that builds international consensus while upholding fundamental rights.[260] The thesis endorses this direction: harmonisation, whether through transatlantic agreements, the OECD AI principles, or future U.N. guidelines, can prevent regulatory arbitrage and ensure that people's rights are protected regardless of where an AI system is developed or deployed. In sum, governing AI-powered decision-making requires more than isolated national laws; it demands a coherent, rights-centric approach that keeps pace with technological change. By reaffirming the primacy of human rights and embedding ethical design and accountability into AI systems, societies can reap the benefits of innovation while safeguarding individuals against undue harm. This conclusion, therefore, closes with a confident call for **legislative reform and further research** aimed at building a unified, principled, and future-proof framework for AI governance, one that secures public trust and upholds the rights of all in the age of algorithms.

[260] HLEG on AI, Ethics Guidelines for Trustworthy AI (n 211)

Bibliography

Primary Sources

European Union Legislation

Charter of Fundamental Rights of the European Union [2012] OJ C 326.

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1.

United States Legislation

Algorithmic Accountability Act of 2019, HR 2231, 116th Congress (2019).

Algorithmic Accountability Act of 2022, HR 6580, 117th Congress (2022).

Algorithmic Accountability Act of 2023, HR 5628, 118th Congress (2023).

Algorithmic Accountability Act of 2025, S 2164, 119th Congress (2025).

Colorado Senate Bill 21-169 (Restrict Insurers' Use of External Consumer Data and Information) (2021).

Case Law

Case C-203/22 *Dun & Bradstreet Austria GmbH v CRIF GmbH* EU:C:2025:XYZ, Judgment of 27 February 2025.

Case C-293/12 and C-594/12 Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others [2014] ECLI:EU:C:2014:238.

Case C-362/14 Maximillian Schrems v Data Protection Commissioner [2015] ECLI:EU:C:2015:650.

Case C-634/21 *OQ v Land Hessen* EU:C:2023:XYZ, Judgment of 7 December 2023 (SCHUFA Holding AG).

United States Official Documents

Algorithmic Accountability Act of 2023, Discussion Draft, 118th Congress (Sen Ron Wyden, 2023) https://www.wyden.senate.gov/imo/media/doc/algorithmic_accountability_act_text.pdf

Executive Order No 13859, Maintaining American Leadership in Artificial Intelligence (11 February 2019) 84 Fed Reg 3967.

Executive Order No 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (30 October 2023) 88 Fed Reg 73589.

Office of Management and Budget, *Guidance for Regulation of Artificial Intelligence Applications* Memorandum M-21-6 (17 November 2020).

Wyden, Booker and Clarke, 'Wyden, Booker and Clarke Introduce Algorithmic Accountability Act of 2022 To Require New Transparency and Accountability for Automated Decision Systems' (Press Release, US Senate, 3 February 2022) https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems

European Union Official Documents

Marco Almada, *Law & Compliance in AI Security & Data Protection, Training Curriculum* (European Data Protection Board, Support Pool of Experts programme, December 2024).

High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (European Commission, 8 April 2019) https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission, *AI Act* (EU Digital Strategy, 30 July 2024) (image 'AI pyramid of risks according to AI Act').

Canadian Official Documents

Office of the Privacy Commissioner of Canada, *Artificial Intelligence Consultation – Final Submissions* (OPC, November 2020) https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/pol-ai_202011/

Office of the Privacy Commissioner of Canada, *A Regulatory Framework for AI: Recommendations for PIPEDA Reform* (OPC, November 2020).

Office of the Privacy Commissioner of Canada, *Policy Proposals for PIPEDA Reform: Automated Decision-Making and Accountability Module* (OPC, November 2020).

Reports

Box S, *The Road Ahead for U.S. Policy on Artificial Intelligence* (Background Paper No 31, ORF America, 31 March 2025) https://orfamerica.org/newresearch/us-policy-on-artificial-intelligence

European Union Agency for Fundamental Rights, *GDPR in practice – Experiences of data protection authorities* (11 June 2024) https://fra.europa.eu/en/publication/2024/gdpr-experiences-data-protection-authorities

Harris L, *Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress*(Congressional Research Service, 4 June 2025) https://www.congress.gov/crs_external_products/R/PDF/R48555/R48555.2.pdf

Monteleone S and Puccio L, *The CJEU's Schrems Ruling on the Safe Harbour Decision* (European Parliamentary Research Service, October 2015) https://www.europarl.europa.eu/RegData/etudes/ATAG/2015/569050/EPRS_ATA(2015)569050_EN.pdf

Office of the Privacy Commissioner of Canada, *OPC Leadership on AI and Privacy* (OPC, 6 May 2025) https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/leadership-ai/

Privacy International, *Challenge against Clearview AI in Europe* (Privacy International Legal Action, 27 May 2021) https://privacyinternational.org/legal-action/challenge-against-clearview-ai-europe

The White House, *America's AI Action Plan* (July 2025) https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

Senator Ron Wyden, *Algorithmic Accountability Act of 2023: Summary* (May 2023) https://www.wyden.senate.gov/imo/media/doc/algorithmic_accountability_act_of_2023_summary.pdf

International Instruments

OECD, *Recommendation of the Council on Artificial Intelligence* (adopted May 2024; revising 2019 OECD AI Principles).

Recommendation on the Ethics of Artificial Intelligence (UNESCO General Conference, adopted November 2021).

Secondary Sources

Books

Eubanks V, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St Martin's Press 2018).

Hustinx P, 'EU Data Protection Law: The Review of Directive 95/46/EC and the General Data Protection Regulation' in Paul Craig and Gráinne de Búrca (eds), *New Technologies and EU Law* (Oxford University Press 2017) https://doi.org/10.1093/acprof:oso/9780198807216.003.0005

Vrabec HU, *Data Subject Rights under the GDPR* (OUP 2021) https://doi.org/10.1093/oso/9780198868422.001.0001

Journal Articles

Batlle S and van Waeyenberge A, 'EU–US Data Privacy Framework: A First Legal Assessment' (2024) 15(1) *European Journal of Risk Regulation* 191

Custers B and others, '*Assessing the Legal and Ethical Impact of Data Reuse: Developing a Tool for Data Reuse Impact Assessments (DRIA)'* (2019) 5 European Data Protection Law Review 317 https://doi.org/10.21552/edpl/2019/3/7

Goodman B and Flaxman S, '*European Union regulations on algorithmic decision-making and a "right to explanation*"' (2017) 38 AI Magazine 50.

Gstrein OJ and Zwitter AJ, '*Extraterritorial Application of the GDPR: Promoting European Values or Power?*' (2021) 10 *Internet Policy Review* https://doi.org/10.14763/2021.3.1576

Kuziemski M and Pałka P, *AI Governance Post-GDPR: Lessons Learned and the Road Ahead* (European University Institute, 2019) School of Transnational Governance Policy Brief Series, Issue 2019/07 https://doi.org/10.2870/470055

Langenkamp M, Costa A and Cheung C, *Hiring Fairly in the Age of Algorithms* (2020) arXiv, DOI: 10.48550/arXiv.2004.07132 https://doi.org/10.48550/arXiv.2004.07132

Lévesque M, 'Smoke and Mirrors? AI Regulation and Corporate Power' (2024) *University of Illinois Journal of Law, Technology & Policy* (2025) Issue 1 https://ssrn.com/abstract=4736131

Malgieri G and Comandé G, '*Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation'* (2017) 7 International Data Privacy Law 243.

Mittelstadt B, '*Principles Alone Cannot Guarantee Ethical AI*' (2019) *Nature Machine Intelligence*https://ssrn.com/abstract=3391293

Mökander J, Juneja P, Watson DS and Floridi L, '*The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?*' (2022) 32 *Minds and Machines* 751 https://doi.org/10.1007/s11023-022-09612-y

Paal B, 'Article 22 GDPR: Credit Scoring Before the CJEU' (2023) 4 *Global Privacy Law Review* 127.

Phillips M, 'International Data-Sharing Norms: From the OECD to the General Data Protection Regulation (GDPR)' (2018) 137 *Human Genetics* 575 https://doi.org/10.1007/s00439-018-1919-7

van Haperen O, 'GDPR Enforcement Beyond EU-Borders — The Dutch Data Protection Authority's Fine on Clearview AI and the Future of AI Regulation & Enforcement' (2025) 26 *Computer Law Review International* 10 https://doi.org/10.9785/cri-2025-260103 accessed 11 August 2025

Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 76 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469 accessed 18 August 2025.

Zarsky T, '*Incompatible: The GDPR in the Age of Big Data'* (2017) 47 The Seton Hall Law Review 2 <https://scholarship.shu.edu/cgi/viewcontent.cgi?article=1606&context=shlr>

Conference Paper

Jung WK and others, 'Privacy and Data Protection Regulations for AI Using Publicly Available Data: Clearview AI Case' in *Proceedings of the 17th International Conference on Theory and Practice of Electronic Governance* (ACM 2024) 48 https://doi.org/10.1145/3680127.3680200

Online Resources

European Centre for Not-for-Profit Law, Liberties and European Civic Forum, 'Packed with loopholes: why the AI Act fails to protect civic space and the rule of law' (ECNL, 3 April 2024) https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law

Hoffmann M, 'AI Safety under the EU AI Code of Practice — A New Global Standard?' (Center for Security and Emerging Technology, 30 July 2025) https://cset.georgetown.edu/article/eu-ai-code-safety/

Jasserand C, 'Clearview AI: Illegally Collecting and Selling Our Faces in Total Impunity? (Part II)' (CiTiP Blog, KU Leuven, 5 May 2022) https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-ii/

Malgieri G, '"Just" Algorithms: AI Justification (beyond explanation) in the GDPR' (Blog, 14 December 2020) https://www.gianclaudiomalgieri.eu/2020/12/14/just-algorithms/
Mazur J, Novelli C and Choińska Z, 'Should Data Protection Authorities Enforce the AI Act? Lessons from EU-wide Enforcement Data' (12 June 2025) SSRN https://ssrn.com/abstract=5290513 or http://dx.doi.org/10.2139/ssrn.5290513

HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition (Electronic Privacy Information Center, 12 January 2021) https://www.epic.org/hirevue-facing-ftc-complaint-from-epic-halts-use-of-facial-recognition/