VRIJE UNIVERSITEIT BRUSSEL

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Laws in de rechten

Human Oversight in EU Data Protection Law: A Study of the GDPR and the AI Act

Michaël THOMAS
0556985

Academic year 2024-2025

Promotor: Gloria GONZALEZ FUSTER

Jury: Rocco Bellanova

Recht & Criminologie

DECLARATION OF ORIGINALITY

The signed declaration of originality and publication is an integral component of the written work (e.g. bachelor or master thesis) submitted by the student.

With my signature I confirm that:

- I am the sole author of the written work here enclosed[1];
- I have compiled the written work in my own words;
- I have not committed plagiarism as defined in article 118 of the Teaching and examination regulations of the VUB; of which the most common forms of plagiarism are (non-exhaustive list):
  - form 1: to copy paragraphs/sentences of other authors with a reference, but without the use of appropriate quotation marks when needed;
  - form 2: to copy paragraphs/sentences of other authors, whether literally or not, without any sources;
  - form 3: to refer to primary sources where both text and references, whether literally or not, are copied from a non-referenced secondary source;
  - form 4: to copy paragraphs/sentences of other authors, whether with a reference or not, with only minimal or misleading adaptions.
  - form 5: to copy own previous work without proper citation (self-plagiarism).
  - form 6: to generate or substantially modify text fragments with software tools, without explicit mention of this
- I have referenced fully in the text and in the reference list any internet sources, published or unpublished works from which I have quoted or drawn reference;
- I have indicated clearly all text from which I have quoted literally;
- I have documented all methods, data and processes truthfully;
- I have not manipulated any data;
- I have mentioned all persons and organisations who were significant facilitators of the work, thus all the work submitted for assessment is my own unaided work except where I have explicitly stated otherwise;
- this work nor part of it has been submitted for a degree to any other programme, university or institution;
- I am aware that the work will be screened for plagiarism and the misuse of generative AI;
- I am aware that I need to disclose the use of generative AI in accordance with the relevant guidelines;
- I will submit all original research material immediately at the Dean's office when it is requested;
- I am aware that it is my responsibility to check when I am called for a hearing and to be available during the hearing period;
- I am aware of the disciplinary sanctions in case of irregularities as described in article 118 of the Teaching and examination regulations of the VUB;
- (Only for master theses) I agree to the publication of my master thesis. If I do not agree, I will attach the 'consent form for publicising master thesis' form together with my master thesis**.** (to be found on the VUB website).

Student's surname, first name: Thomas Michaël **Date:** 14/08/2025

Signature: Thomas

[1] For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Consent form for publicising Master's thesis

Student: Thomas Michaël
Enrolment number : 0556985
Study programme : Master of Laws in de Rechten
Academic year: 2024-2025

Master's thesis
Title: Human Oversight in EU Data Protection Law: A Study of the GDPR and the AI Act
Supervisor: Prof. dr. Gloria González Fuster

Any Master's thesis for which the student has obtained a credit, and for which no non-disclosure agreement (NDA) was drawn up, can be included at no charge in the Vubis catalogue of the central university library as long as the student has given their prior explicit consent.

In the context of the possibility of making their Master's thesis available free of charge, the student chooses one of the following options:

- ☒ OPEN ACCESS: worldwide access to the full text of the Master's thesis
- ☐ ONLY FROM THE CAMPUS: access to the full text of the Master's thesis is only possible from the VUB network
- ☐ EMBARGO FOLLOWED BY OPEN ACCESS: worldwide access to the full text of the Master's thesis only after a specific date, namely .....
- ☐ EMBARGO FOLLOWED BY ACCESS ONLY FROM THE CAMPUS: access to the full text of the Master's thesis only from the campus and only after a specific date, namely .....
- ☐ FULL TEXT NEVER ACCESSIBLE: no access to the full text of the Master's thesis
- ☐ NO PERMISSION to make accessible

The supervisor confirms acknowledgement of the intention of the student to make the Master's thesis available in the Vubis catalogue of the central university library.

Date:
Signature of supervisor:

This document will be included in the Master's thesis. Any student who fails to include the document in their Master's thesis and/or has failed to indicate a choice and/or has failed to sign the document and/or has failed to inform the supervisor will be considered as not having granted permission to make their thesis public; in that event, the Master's thesis will only be archived and will not be accessible to the public.

Drawn up at Etterbeek ..... on 14 August 2025

Signature of student 

Acknowledgements

This thesis marks the conclusion of my studies in law at the Vrije Universiteit Brussel. While it represents an academic milestone, it is above all a personal one. It is the result of years of learning, reflection, and perseverance, shaped by both enriching experiences and profound challenges. I wish to express my sincere gratitude to those who have supported me throughout this journey.

First and foremost, I thank my parents. Their unwavering support, financially and above all morally, was the foundation that allowed me to move forward in moments of uncertainty. At times when the path was anything but smooth and the obstacles unlike those many students will ever have to face, they stood firmly beside me. Part of this journey included a serious illness that forced me to pause, recover, and rebuild. What followed was not always easy, but Perseverantia Vincit! I carry that lesson with me always. Their belief in me helped sustain my determination when it mattered most. If I have reached this point, it is in no small part due to their love and presence.

I would also like to express my deepest thanks to my former colleagues and now friends at the Brussels Privacy Hub: Michaël, Alessandro, and Barbara. Thank you for your presence during this process and far beyond. Your encouragement, thoughtful advice, and the light you brought in more difficult times have meant more than I can put into words. I am equally grateful for the opportunity I had to work within the Brussels Privacy Hub, where I learned and grew both intellectually and personally, surrounded by people whose work I respect and whose friendship I now treasure. I am especially thankful to Professor Sophie Stalla-Bourdillon, whose trust, mentorship, and generosity of spirit opened doors I had not expected.

I also wish to thank Professor González Fuster, whose guidance during the thesis process was essential. Her course in my first master's year was my first introduction to data protection law and helped ignite the passion that would come to shape much of my academic path and professional interests.

I would also like to thank Professor Gori, whom I had the pleasure of being taught by in the first semester of my final year. His passionate and engaging teaching, combined with a nuanced and practically grounded perspective, helped shape the way I think about complex legal questions, especially in relation to the AI Act. I am particularly grateful for his availability and generosity, always offering support far beyond what was required.

Finally, I thank all those who have walked alongside me, offered encouragement, challenged my thinking, and stood by me during these transformative years. This thesis carries my name, but it was made possible through the care, kindness, and insight of many others.

As I prepare to turn the page, I know I will miss the people, ideas, and spirit of the VUB. In particular, its strong emphasis on free and critical thinking has shaped the way I approach both legal questions and the world around me. At the same time, I look ahead with anticipation, grateful for what I have learned and ready to begin what comes next.

Ex animo grates persolvo,

Michaël Thomas

14 August 2025

DISCLAIMER

I have made limited use of linguistic Artificial Intelligence tools in the preparation of this thesis. These tools were employed exclusively to assist in rewriting and refining text that I originally drafted, including grammar, vocabulary, and stylistic improvements. I also used AI to generate preliminary suggestions for the overall structure and to support the ideation process, particularly by testing and evaluating the strength of legal arguments and critical positions. At all times, I exercised full control over the content, and I take full responsibility for the final result. Where content has been paraphrased, this has been done with care to preserve the original meaning and intent of the cited primary and secondary sources, which are properly referenced in the footnotes and bibliography.[1]

[1] This approach aligns with the policy position of the Vrije Universiteit Brussel regarding the responsible use of (generative) Artificial Intelligence in research processes. See Vrije Universiteit Brussel, 'Policy and Structure' https://www.vub.be/en/our-research/policy-and-structure.

Table of Contents

List of Abbreviations

ADM: Automated decision-making

AEPD: Agencia Española de Protección de Datos

AI: Artificial Intelligence

AG: Advocate General

AP: Autoriteit Persoonsgegevens

CFR: Charter of Fundamental Rights of the European Union

CJEU: Court of Justice of the European Union

CNIL: Commission nationale de l'informatique et des libertés

DPA: Data Protection Authority

DPD: Data Protection Directive

DPIA: Data Protection Impact Assessment

EDPB: European Data Protection Board

EDPS: European Data Protection Supervisor

EU: European Union

FRIA: Fundamental Rights Impact Assessment

GBA: Gegevensbeschermingsautoriteit

GDPR: General Data Protection Regulation

LED: Law Enforcement Directive

SCHUFA: Schutzgemeinschaft für allgemeine Kreditsicherung

TEU: Treaty on European Union

TFEU: Treaty on the Functioning of the European Union

WP29: Article 29 Working Party

1 Introduction

In recent decades, concerns about automated decision-making have played a central role in the development of data protection law. Since the 1970s, legislators have recognised the risks posed by decisions made without human involvement, especially when such decisions evaluate or define aspects of a person's behaviour or identity. Early national legislation, such as the French Data Protection Act of 1978, introduced strict prohibitions on profiling in judicial, administrative, and private contexts, and guaranteed individuals the right to understand and challenge decisions made about them.[2] This concern was later codified at the European level in Article 15 of the 1995 Data Protection Directive, which required Member States to ensure that individuals were not subject to decisions based solely on automated processing that significantly affected them.[3] However, the Directive permitted divergent national interpretations. Some countries, like Belgium, adopted a strict prohibition model, while others, like Sweden, allowed greater leeway for data controllers.[4] Article 22 of the General Data Protection Regulation was introduced in 2016 to harmonise this fragmented legal landscape by setting out a uniform standard for ADM and establishing a more structured approach to human intervention.[5]

This historical foundation is increasingly relevant as AI systems gain prominence across the public and private sectors. From credit scoring and hiring to welfare allocation and fraud detection, algorithmic systems are now used to make or inform decisions that have profound consequences for individuals.[6] As the scale and complexity of these technologies grow, so do concerns about their fairness, opacity, and accountability. ADM can produce outcomes that are difficult to trace or contest, often reinforcing existing inequalities or structural biases.[7] The introduction of increasingly powerful AI tools has amplified these concerns, prompting lawmakers and regulators to reconsider how human oversight

[2] Article 47 Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.
[3] 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 ('DPD').
[4] Wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens, BS 3 februari 1999 (Belgian Privacy Act); Sebastião Barros Vale, 'GDPR and the AI Act Interplay: Lessons from FPF's ADM Case-Law Report - Future of Privacy Forum' (*Future of Privacy Forum*, 2022) 7 <https://fpf.org/blog/gdpr-and-the-ai-act-interplay-lessons-from-fpfs-adm-case-law-report/> accessed 5 August 2025.
[5] 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 [2016] OJ L119/1 ('GDPR').
[6] High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI | Shaping Europe's Digital Future' (2019) 9 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 22 December 2024.
[7] ibid 18; 'EDPB-EDPS Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' para 59.

can be effectively embedded into automated systems. Similar debates and regulatory initiatives are emerging beyond the EU, with jurisdictions such as the United States, Canada and several Asian countries exploring frameworks for algorithmic accountability and human oversight.[8] While the challenge is global, the EU has taken the lead by adopting the most comprehensive and protective approach to AI regulation to date, exemplified by the recently adopted Artificial Intelligence Act, which introduces dedicated rules for high-risk AI applications, including specific requirements for human oversight.[9]

The human oversight requirement is by no means unique to the GDPR or AI Act. A 2022 policy survey identified over 40 policy instruments from across the world that mandate or encourage human oversight in the context of public-sector algorithms.[10] Within EU law, legally binding oversight obligations can also be found in instruments such as the Digital Services Act and the Law Enforcement Directive.[11] However, these frameworks fall outside the scope of this thesis. The GDPR remains the central legal instrument governing data protection in the EU, while the DSA addresses broader content moderation and platform regulation without focusing on personal data governance. This thesis therefore concentrates on the evolving relationship between Article 22 of the GDPR and Article 14 of the AI Act. Article 22 of the GDPR sets limits on decisions based solely on automated processing that produce legal or similarly significant effects, whereas Article 14 of the AI Act introduces operational requirements for human oversight in high-risk AI systems. By examining how these two provisions interact, the analysis assesses whether the AI Act addresses the limitations of the GDPR and whether it offers a more robust legal foundation for ensuring meaningful human involvement in automated decision-making processes.

[8] Kostiantyn Ponomarov, 'Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview' (25 August 2025) <https://legalnodes.com/article/global-ai-regulations-tracker> accessed 21 September 2025; 'AI Watch: Global Regulatory Tracker - China | White & Case LLP' (29 May 2025) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china> accessed 21 September 2025.

[9] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L2024/1689 ('AI Act').

[10] Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 Computer Law & Security Review 105681, 17–18 <https://www.sciencedirect.com/science/article/pii/S0267364922000292> accessed 28 April 2024.

[11] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [2022] OJ L227/1 ('DSA'); Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision [2016] OJ L119/89 ('LED').

To guide this analysis, the thesis is structured around one main research question: 'How does the AI Act reinforce, or diverge from the GDPR's framework for human oversight?' This overarching question is explored through three sub-questions, each of which is addressed in a dedicated chapter. Chapter 2 lays the conceptual foundations by clarifying key notions, such as artificial intelligence, automated decision-making, and human oversight. Chapter 3 responds to the first sub-question: 'What mechanisms currently support human oversight under the GDPR, and what limitations exist?' It provides a legal analysis of the GDPR's provisions and identifies areas of structural weakness. Chapter 4 addresses the second sub-question: 'What new provisions does the AI Act introduce specifically for human oversight?' It analyses Article 14 AI Act and related obligations, paying attention to their legal structure, objectives, and operational features. Finally, Chapter 5 answers the third sub-question: 'To what extent does the AI Act address the limitations identified in the GDPR's approach to human oversight, and what gaps remain?' It evaluates whether the AI Act complements the GDPR in ways that improve both individual protection and institutional accountability, or whether key limitations persist. This sequence also positions the thesis within the current academic debate. Much of the existing literature questions whether Article 22 of the GDPR provides an effective model of human oversight, pointing to its narrow scope and reliance on individual rights. More recent work examines whether the AI Act's risk-based framework can overcome those weaknesses and establish stronger, design-stage obligations. By systematically comparing the two regimes and assessing their interaction, this thesis contributes to that debate by showing how the AI Act both responds to and leaves open the gaps identified in the GDPR. It therefore offers a bridge between these two strands of scholarship and clarifies the extent to which EU law is shifting from reactive remedies toward a more structural conception of human oversight.

In terms of methodology, this research follows a doctrinal legal approach, grounded in the close analysis of EU legislation, case law, and regulatory guidance. It examines primary legal texts, such as the GDPR and the AI Act, alongside preparatory documents, recitals, and related instruments where relevant. The thesis also draws on academic commentary, institutional reports, and decisions from data protection authorities and courts to identify interpretative trends and practical challenges. Particular attention is paid to sources such as the European Data Protection Board's guidance on automated decision-making and profiling, recent judgments from the Court of Justice of the European Union, and the ongoing debate around the implementation of human oversight in high-risk AI systems. The aim is to provide a comprehensive and critical account of the legal framework, grounded in both theory and practice. Where relevant, AI-assisted translation tools were used to analyse foreign-language DPA decisions and national case law, with all interpretations cross-checked against contextual and doctrinal understanding.

2 Terminological Clarifications

Before analysing the legal frameworks and oversight mechanisms in the GDPR and the AI Act, achieving terminological clarity regarding three pivotal terms supporting the thesis is important: artificial intelligence, automated decision-making, and human oversight. Furthermore, these concepts unveil vital contrasts amid the standardising and functional difficulties of how we govern algorithmic systems. Clarifying these terms provides the necessary analytical grounding for the sub-research questions that follow.

2.1 Artificial Intelligence

Artificial intelligence is a notoriously difficult concept to define with precision. In its broadest sense, AI refers to a set of techniques and systems capable of performing tasks that would normally require human intelligence, such as learning, reasoning, pattern recognition, or decision-making. However, what qualifies as "intelligent" behaviour varies across disciplines and contexts, making AI an inherently polysemic term.[12]

Scholars have long noted that AI is best understood as an umbrella term rather than a fixed technological category.[13] AI encompasses a heterogeneous set of computational techniques aimed at simulating cognitive functions, ranging from expert systems to machine learning and neural networks.[14] AI can also be described as “a moving target,” subject to continuous shifts as technologies evolve and previously novel methods become ordinary.[15] The term also carries social and political weight, often used rhetorically to evoke innovation or risk.

[12] See for example AM Turing, ‘Computing Machinery and Intelligence’ (1950) LIX Mind 433 <https://doi.org/10.1093/mind/LIX.236.433>; Stuart Russel and Peter Norvig, *Artificial Intelligence A Modern Approach* (Fourth Edition Global Edition, Pearson 2021); Pamela McCorduck and Cli Cfe, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (2nd edition, A K Peters/CRC Press 2004).

[13] Dr Gizem Gültekin-Várkonyi, ‘AI Literacy for Legal AI Systems: A Practical Approach’ (Social Science Research Network, 20 May 2025) 1–2 <https://papers.ssrn.com/abstract=5309725> accessed 21 July 2025; Hannah Ruschemeier, ‘AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal’ (2023) 23 ERA Forum 361, 364–365 <https://doi.org/10.1007/s12027-022-00725-6> accessed 26 July 2025.

[14] Arun Rai, Panos Constantinides and Saonee Sarker, ‘Next Generation Digital Platforms : Toward Human-AI Hybrids’ (2019) 43 MIS Quarterly iii, <https://misq.org/misq/downloads/> accessed 21 July 2025.

[15] Recital 4 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L2024/1689 (‘AI Act’).

Given this conceptual fluidity, the EU has had to operationalise a legal definition of AI that is both sufficiently precise for regulation and broad enough to future-proof the legislation. The European Commission's 2021 draft of the AI Act originally adopted a broad definition:

> *"Artificial intelligence system (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with."*[16]

This definition was heavily criticised by stakeholders, who feared it would apply to a wide range of software systems with minimal "intelligent" behaviour.[17] After extensive negotiations and lobbying, the final text of the AI Act adopted a slightly more restricted definition. It now defines an AI system as:

> *"a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."*[18]

The revised definition narrows the scope in three ways. First, it removes direct references to specific techniques, instead requiring the system to infer how to produce outputs based on input. Second, it adds that the system may exhibit adaptiveness after deployment, distinguishing AI from static rule-based software. Third, it explicitly includes systems with varying levels of autonomy, clarifying that both fully and partially automated tools can fall within the Act's scope.

This legal definition does not aim to capture the full philosophical or technical richness of the concept of AI. Instead, it serves a regulatory function, anchoring obligations and risk-based classifications within the AI Act's framework. For the purposes of this thesis, this legal definition will be used as a point of reference when analysing how human oversight duties apply to AI systems under EU law.

---

[16] European Commission, 'Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM(2021) 206 final, 2021/0106(COD).

[17] Stakeholders criticised the initial definition of AI in the 2021 draft Act as overly broad, warning that it would capture systems with minimal intelligent behaviour, such as simple search, sorting, and routing algorithms, and thus subject conventional software to disproportionate obligations. Among others, the Big Data Value Association, AmCham EU, and the European Digital SME Alliance called for narrowing the definition and reducing uncertainty about its scope: Tambiama Madiega, *Artificial Intelligence Act* (EU Legislation in Progress Briefing, European Parliamentary Research Service, PE 698.792, September 2024) 4–5.

[18] Article 3(1) AI Act.

2.2 Automated Decision-Making

ADM refers to processes through which decisions are generated by computational systems without substantive human involvement.[19] It typically involves collecting input data, analysing it using programmed logic or statistical models, and producing an output that determines or influences an individual's status, treatment, or access to services.[20] While often associated with recent advances in artificial intelligence, ADM also encompasses more traditional techniques such as decision trees or expert systems that mechanically apply pre-established rules. The common thread is that the determination is not made through human judgment, but through algorithmic operations acting on data.

While ADM is frequently discussed in the context of artificial intelligence, the two concepts are not synonymous. ADM refers to algorithmically driven processes that play a decisive role in producing outcomes.[21] AI, by contrast, is a broader technological category encompassing a variety of computational approaches, some of which do not produce decisions at all. For instance, AI may be used to generate images, simulate environments, or optimise systems without directly determining how a person is treated. Conversely, ADM can be implemented using simple rule-based tools that do not qualify as AI under evolving legal definitions, including that of the AI Act. This means that not all ADM qualifies as AI, and not all AI results in ADM. However, many systems do fall within the overlap of both categories, and these AI-driven decision-making systems form the core focus of this thesis.[22]

ADM often overlaps with profiling, but the two are also not equivalent. Profiling refers to the automated evaluation of personal aspects and data in order to predict or classify individuals.[23] While profiling may inform ADM, it does not necessarily lead to a decision.[24] Similarly, ADM may be based on static criteria that are not generated through profiling.

[19] 'Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration' (European Law Institute) 16–17; Teresa Rodríguez de las Heras Ballell, 'Guiding Principles for Automated Decision-Making in the EU' (2022) ELI Innovation Paper 8.
[20] ibid.
[21] Teresa Rodríguez de las Heras Ballell (n 19) 9.
[22] This overlap is discussed in Section 5.1.2.
[23] Article 4(4) GDPR.
[24] Valeria Ferraris and others, 'Defining Profiling' (Social Science Research Network, 11 December 2013) 3–4 <https://papers.ssrn.com/abstract=2366564> accessed 22 July 2025.

2.3 Human Oversight

2.3.1 Concept

The principle of human oversight is commonly promoted as a core safeguard against the risks associated with ADM.[25] As a principle, it embodies the idea that human judgment, contextual sensitivity, and ethical reasoning can serve as a counterbalance to the limitations of algorithmic systems.[26] Oversight by humans is intended to ensure that decisions affecting individuals are not left solely to automated processes, thereby protecting fundamental rights, upholding fairness, fostering accountability, and securing respect for human autonomy.[27] Such oversight reflects a commitment to human-centric design: AI systems should be developed to augment, complement, and empower human cognitive, social, and cultural capacities, rather than unjustifiably subordinate, coerce, or manipulate individuals.[28] At its core, human-centric AI reflects the ethical premise that human beings possess a unique moral and legal status, one that technological development must not undermine.[29] While human oversight tends to be reactive, functioning as a procedural safeguard to mitigate risks once AI systems are in operation, human-centric design takes a more proactive stance.[30] It seeks to embed human values and societal needs into the very architecture of AI systems from the outset.[31] Underlying both the reactive and proactive dimensions of oversight is the broader normative idea of human agency, understood as the capacity of individuals to understand, influence, and contest decisions that significantly affect their lives.[32] This notion of agency informs the protective function that oversight aims to fulfil, linking it to deeper questions of autonomy, accountability, and democratic control.

2.3.2 Objectives

To better understand the regulatory aims behind human oversight obligations, it is helpful to distinguish between the different goals they may serve. Fink proposes a tripartite typology of oversight

[25] Recital 71 GDPR.
[26] Johann Laux, 'Institutionalised Distrust and Human Oversight of Artificial Intelligence: Towards a Democratic Design of AI Governance under the European Union AI Act' (2024) 39 AI & SOCIETY 2853, 2855 <https://doi.org/10.1007/s00146-023-01777-z> accessed 25 June 2025.
[27] High-Level Expert Group on AI (n 6) 12.
[28] ibid; European Commission, 'White Paper on Artificial Intelligence – A European Approach to Excellence and Trust' COM(2020) 65 final 3.
[29] Dominika Harasimiuk and Tomasz Braun, *Regulating Artificial Intelligence: Binary Ethics and the Law* (Routledge 2021) 49–50.
[30] Lena Enqvist, '"Human Oversight" in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15 Law, Innovation and Technology 508, 511 <https://doi.org/10.1080/17579961.2023.2245683> accessed 22 December 2024.
[31] ibid.
[32] High-Level Expert Group on AI (n 6) 15–16.

objectives: output-oriented goals, process-oriented goals, and accountability-oriented goals.[33] Output-oriented oversight focuses on detecting and correcting erroneous or harmful outcomes, ensuring that algorithmic outputs do not go unchecked.[34] Process-oriented oversight aims to safeguard procedural fairness, transparency, and human agency during the decision-making process itself.[35] Accountability-oriented oversight is concerned with clarifying who bears legal and ethical responsibility when harm occurs.[36] A practical illustration is Tesla's June 2025 pilot of autonomous "robotaxis" in Austin, which still include a "safety driver" in the passenger seat.[37] This safety driver can intervene to correct unsafe driving outputs, thereby fulfilling an output-oriented function, while also serving as the legally responsible human counterpart should anything go wrong, which addresses accountability-oriented goals. This demonstrates that a single oversight arrangement may simultaneously fulfil multiple goals.

2.3.3 Governance Models

In addition to clarifying the normative aims of human oversight, it is also useful to distinguish the different operational forms such oversight can take. A commonly referenced typology includes four conceptual models[38]: human-in-the-loop, human-on-the-loop, human-out-of-the-loop, and human-back-in-control.[39] These models represent varying intervention frameworks, differentiated by the timing of the intervention, the degree of human involvement, and the mechanisms by which such involvement is triggered.[40] In the human-in-the-loop model, the output of an AI system does not take effect unless it is first reviewed and validated by a human, such as when the rejection of a social security application must be confirmed by a human agent.[41] The human-out-of-the-loop model permits

[33] Melanie Fink, 'Human Oversight under Article 14 of the EU AI Act' (Social Science Research Network, 14 February 2025) 7 <https://papers.ssrn.com/abstract=5147196> accessed 25 June 2025.
[34] ibid 7–8, referring to the 1983 Petrov incident where a Soviet officer overrode a false nuclear alert, and the case of Somali-owned stores in Seattle flagged by an algorithm for fraud due to culturally specific transaction patterns.
[35] ibid 8, noting how human involvement supports procedural rights, public sector discretion, and democratic legitimacy, including the example of preserving discretion in frontline decision-making.
[36] ibid 8–9, discussing how human presence in self-driving cars serves to assign liability, including Tesla's design transferring control back to the driver just before impact to offload responsibility.
[37] Brad Anderson, 'Tesla Says It's Driverless But Someone's Always Watching' (*Carscoops*, 23 June 2025) <https://www.carscoops.com/2025/06/tesla-robotaxis-arrive-in-austin-but-its-using-safety-drivers/> accessed 22 July 2025.
[38] These models constitute a non-exhaustive typology of intervention frameworks and serve primarily as conceptual tools, as clarified in the White paper on Artificial Intelligence by the EC. They do not represent a legally codified or comprehensive classification.
[39] Guillermo Lazcoz and Paul de Hert, 'Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities' (2023) 50 Computer Law & Security Review 105833, 7–8 <https://www.sciencedirect.com/science/article/pii/S0267364923000432> accessed 15 March 2025.
[40] European Commission (n 28) 21.
[41] Lazcoz and de Hert (n 39) 7; Rebecca Crootof, Margot Kaminski and W. Price II, 'Humans in the Loop' (2023) 76 Vanderbilt Law Review 429, 440–442 <https://scholarship.law.vanderbilt.edu/vlr/vol76/iss2/2>.

the system's output to take immediate effect, but allows for subsequent human intervention, for instance in cases of automatic credit card rejections where appeal to a human reviewer remains possible.[42] In the human-on-the-loop model, a human monitors the system in real-time and has the capacity to intervene or deactivate it if necessary, such as by pressing a stop button in the context of autonomous vehicles.[43] Finally, the human-back-in-control model refers to systems that include predefined operational thresholds or conditions under which control is automatically transferred back to a human operator, as may occur in low visibility scenarios encountered by driverless cars.[44] These models are conceptual governance tools that can assist in evaluating the credibility and effectiveness of human oversight mechanisms. They serve as a useful analytical lens for understanding how oversight is structured and implemented across regulatory contexts.

2.3.4 Regulatory Framing

The GDPR, and Article 22 GDPR specifically, positions human involvement as an essential protective mechanism intended to ensure fair, transparent, and accountable decision-making.[45] This provision reflects a regulatory commitment to the idea that human intervention can mitigate the risks associated with automation by introducing contextual judgment and ethical evaluation. Recital 71 GDPR further highlights this ideal, suggesting that human oversight is capable of preventing discriminatory or otherwise unjust outcomes.[46] Underlying this regulatory choice is a deeply embedded assumption that human oversight is inherently beneficial, capable of mitigating algorithmic risks through contextual judgment and ethical evaluation.[47] The value attributed to human oversight is closely tied to democratic legitimacy, accountability, and respect for fundamental rights.[48]

However, this assumption warrants critical reassessment. Both emerging and long-standing doctrinal perspectives and empirical evidence highlight significant flaws and limitations in relying on human intervention as a safeguard.[49] They suggest that the presence of human oversight, as currently conceived, may provide a false sense of security when adopting algorithmic systems.[50]

[42] Lazcoz and de Hert (n 26) 7; in Fink (n 20) 2, this is also called 'human review'.
[43] Lazcoz and de Hert (n 39) 7.
[44] Lazcoz and de Hert (n 26) 7; in Fink (n 20) 2–3, this is also called 'human design'.
[45] Article 22 GDPR
[46] Recital 71 GDPR.
[47] Lazcoz and de Hert (n 39) 2–3; Green (n 10) 1–2.
[48] High-Level Expert Group on AI (n 6) 15–16.
[49] These limitations will be explored in section 3.2.
[50] Green (n 10) 9.

3 Current Mechanisms Supporting Human Oversight Under GDPR and Their Limitations

This chapter addresses the first research sub-question: What are the current mechanisms supporting human oversight under the GDPR, and what are their limitations in the context of ADM? It begins by examining the relevant legal provisions, including Article 22 GDPR, the underlying principles of data protection law, and applicable data subject rights. Throughout the analysis, relevant case law is used to illustrate how oversight is interpreted and applied in practice. The chapter then turns to critiques and limitations of the current framework.

3.1 Legal Framework of Human Oversight in the GDPR

The GDPR's legal framework for human oversight in automated decision-making reflects both specific rules and broader normative commitments. At its centre is Article 22 GDPR, which seeks to prevent individuals from being subjected to significant decisions made solely by automated means, except in narrowly defined circumstances. This framework is not confined to Article 22 GDPR alone but is shaped by core principles of data protection law that provide essential context for understanding the scope and limits of human oversight in practice. The following sections examine how Article 22 GDPR structures this protection, the exceptions it allows, and the principles that guide its interpretation and enforcement.

A similar provision exists in Article 11 of the LED, which prohibits solely automated decisions in the context of law enforcement, unless authorised by law and subject to safeguards.[51] However, the GDPR remains the primary legal framework for human oversight in most domains, as the LED applies only to processing by competent authorities for criminal justice purposes.

3.1.1 Scope of Article 22 GDPR

3.1.1.1 The Right Not to Be Subject to Automated Decision-Making

The first paragraph of Article 22 GDPR sets out the primary safeguard against ADM and profiling with significant impact, providing that:

[51] Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision [2016] OJ L119/89 ('LED').

> *"The data subject 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."* [52]

This provision articulates the GDPR's clearest legal barrier to consequential decisions that exclude human judgment and positions human intervention as a key protective mechanism against the risks associated with algorithmic decision-making.[53] There has been considerable debate over whether Article 22(1) GDPR should be understood as a subjective right of the data subject or as an a priori prohibition on automated decision-making.[54] Over time, the prevailing interpretation has come to regard it as a prohibition, applicable unless one of the narrow exceptions in paragraph 2 applies.

This interpretation is reinforced by the EDPB, which has clarified that the term "right" in Article 22(1) GDPR should not be understood as conditional on individual invocation.[55] Instead, the provision imposes a structural obligation on controllers to abstain from fully automated decision-making unless one of the narrow exceptions in paragraph 2 applies. Under this reading, the controller bears the responsibility of justifying the use of automated decision-making up front, rather than awaiting objection from the data subject.[56] The CJEU has echoed this approach in the SCHUFA case, noting that the article lays down a prohibition in principle that does not require the individual to invoke it in order for it to apply.[57]

---

[52] Article 22(1) GDPR.

[53] EDPB/WP29, 'Article 29 Data Protection Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, WP251rev.01' (2018) 20–21 <https://ec.europa.eu/newsroom/article29/items/612053/en> accessed 13 March 2025.

[54] Sebastião Barros Vale and Gabriela Zanfir-Fortuna, 'Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities' (Future of Privacy Forum 2022) 7–8 <https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf> accessed 3 July 2025; Almada, Marco, 'Automated Decision-Making as a Data Protection Issue' [2021] Social Science Research Network 5 <https://www.researchgate.net/publication/350759609_Automated_decision-making_as_a_data_protection_issue> accessed 8 August 2025; Luca Tosoni, 'The Right To Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation' (Social Science Research Network, 14 May 2021) 2–5 <https://papers.ssrn.com/abstract=3845913> accessed 2 July 2025.

[55] The EDPB, European Data Protection Board, is an independent European body that ensures that the General Data Protection Regulation and the Law Enforcement Directive are applied consistently across all EU member states. It issues guidelines, recommendations and best practice to clarify the law and promote common understanding of EU data protection laws. While most EDPB guidance is not formally binding, it represents the collective interpretation of all EU data protection authorities and is highly influential in practice, with courts and regulators regularly referencing EDPB guidelines when making decisions; EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' 19 <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en> accessed 28 June 2025.

[56] Anna Levitina, 'Humans in Automated Decision-Making under the GDPR and AI Act' [2024] Revista CIDOB d'Afers Internacionals 121, 126 <https://cidob.org/en/publications/humans-automated-decision-making-under-gdpr-and-ai-act> accessed 26 July 2025.

[57] Case C-634/21 SCHUFA Holding [2023] ECLI:EU:C:2023:957, para 52.

The CJEU has further underlined that the application of Article 22(1) GDPR depends on the presence of three cumulative elements, namely that there is a qualifying decision, that the decision results solely from automated processing including profiling, and that its outcome has legal effects for the individual or otherwise affects them in a comparably significant way.[58] These criteria set the threshold for when the prohibition comes into play, and their precise meaning will be examined in the subsections that follow.

This prohibitionist reading reflects a deliberate regulatory shift from earlier frameworks, such as Article 15 of the DPD, which permitted divergent national implementations. Belgium, for instance, had already embedded a prohibition model in its national law prior to the GDPR.[59] The Belgian Privacy Act of 1992 explicitly forbade decisions producing legal effects or significantly affecting individuals when based solely on automated processing intended to evaluate aspects of a person's personality.[60] By contrast, Article 22 GDPR was designed to harmonise the fragmented legal landscape across the EU by establishing a uniform baseline: individuals should not be subject to consequential automated decisions unless narrowly defined conditions are met.

3.1.1.2 Solely Automated Processing

The wording of Article 22(1) GDPR reflects a deliberate policy choice to both prohibit automated decisions that bypass meaningful human involvement, while at the same time leaving the provision open-ended through the inclusion of exceptions and interpretative ambiguity. The phrase "solely automated processing" signifies that the decision must be made without any genuine human input for the prohibition to apply. The mere presence of nominal human review will not suffice to remove a decision from the scope of Article 22 GDPR.[61] The EDPB has clarified that for oversight to be meaningful, human involvement must have the capacity to influence the outcome, rather than simply endorse the automated result.[62]

This interpretation has been confirmed in several enforcement decisions and judicial rulings. For instance, the Amsterdam District Court found that Uber's system for deactivating driver accounts was not "solely automated" because a specialized human team assessed the system's fraud signals and

[58] ibid, para 43.
[59] Wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens, BS 3 februari 1999 (Belgian Privacy Act).
[60] Article 12bis Belgian Privacy Act.
[61] EDPB/WP29 (n 53) 20–21; Isak Mendoza and Lee A Bygrave, 'The Right Not to Be Subject to Automated Decisions Based on Profiling' (Social Science Research Network, 8 May 2017) 11 <https://papers.ssrn.com/abstract=2964855> accessed 26 June 2025.
[62] EDPB/WP29 (n 53) 21.

unanimously decided on deactivation based on additional contextual factors.[63] Similarly, the Dutch court reviewing the "e-screener" used in firearm licensing ruled that although the algorithm generated a psychological risk score, the final decision rested with a police officer who could override the recommendation after weighing other elements, such as background checks and personal representations.[64]

In a 2022 case involving Grindr, the Spanish DPA, AEPD, assessed whether the dating app's use of a content moderation algorithm to detect illegal activities triggered Article 22 GDPR. Although the system flagged potential violations, the final decision to remove content or block a user account was taken by human moderators after manual review.[65] The AEPD concluded that this constituted sufficient human involvement to remove the process from the scope of Article 22 GDPR.[66]

These rulings suggest that what matters is not whether a human appears somewhere in the process, but whether that human plays a meaningful and autonomous role in shaping the decision.[67] Factors such as the timing, authority, and contextual understanding of the human actor are crucial. Courts and regulators consistently assess the final stage of the decision-making process when determining whether a decision is solely automated. Even in cases where Article 22 GDPR was deemed inapplicable, courts and supervisory authorities often identified violations of other GDPR provisions.[68] In particular, failures relating to transparency, lawfulness, and broader data protection principles frequently led to enforcement actions, highlighting the relevance of other safeguards discussed in Section 3.1.2 of this thesis.

3.1.1.3 Profiling

Profiling plays a central role in understanding the scope and safeguards of Article 22 GDPR**.** Profiling is defined as any automated processing of personal data aimed at evaluating personal aspects of a natural person, particularly to analyse or predict elements such as performance at work, economic situation, health, personal preferences, reliability, behaviour, location, or movements.[69] This definition highlights three essential elements: profiling involves automated processing, it operates on personal data, and its purpose is to assess or predict personal characteristics.[70] Unlike Article 22 GDPR, which

[63] Rechtbank Amsterdam, Case C/13/692003 / HA RK 20-302, ECLI:NL:RBAMS:2021:1018 (11 March 2021) para 4.19, 4.26 <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2021:1018>.
[64] Rechtbank Den Haag, Case C/09/585239 / HA ZA 19-1221, ECLI:NL:RBDHA:2020:865 (11 February 2020) para 4.26 <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1013>.
[65] Agencia Española de Protección de Datos (AEPD), Procedimiento Nº: E/03624/2021(2021) 9–10 <https://www.aepd.es/es/documento/e-03624-2021.pdf>.
[66] ibid 16.
[67] EDPB/WP29 (n 53) 20–21.
[68] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 29.
[69] Article 4(4) GDPR.
[70] EDPB/WP29 (n 53) 6–7.

focuses on solely automated decisions, profiling under Article 4(4) GDPR does not exclude human involvement entirely.[71] This implies that profiling can still be considered solely automated even when humans are involved, provided the automated processing plays a predominant role. Yet, the threshold for when automation becomes predominant remains difficult to define.[72]

Profiling is not limited to simple classifications based on observable attributes like age or gender; it typically entails a layer of inference or judgement about an individual.[73] Profiling is often a staged process[74]: it begins with data collection, followed by automated analysis to identify correlations, and culminates in applying these correlations to individuals to infer likely traits or behaviours.[75] This means profiling generates new personal data by creating or attributing characteristics to individuals based on statistical relationships and patterns. As the Council of Europe explains, profiling applies group-based profiles to individuals identified through collected data, thus creating new attributes linked to that individual.[76] These attributes may or may not accurately reflect the data subject's actual qualities.[77]

Profiling can take individual or group forms. Individual profiling focuses on a specific person's data to infer or predict behaviour, whereas group profiling identifies patterns in a population and applies them to individuals who fit that pattern.[78] Group profiling may rely on distributive models, where shared attributes define group membership, or non-distributive models, where the group is based on correlated risk factors without shared traits (such as predicted health risks from various unrelated indicators).[79] Both forms of profiling can lead to automated decisions within Article 22's scope if they produce legal effects or similarly significant impacts and lack meaningful human review.

[71] Article 4(4), 22 GDPR.
[72] EDPB/WP29 (n 53) 7.
[73] ibid.
[74] For further discussion of multi-stage profiling, see Reuben Binns and Michael Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR' (2021) 11 International Data Privacy Law 319 <https://doi.org/10.1093/idpl/ipab020> accessed 26 July 2025; Mireille Hildebrandt, 'The Disconnect Between "Upstream" Automation and Legal Protection Against Automated Decision Making' (*Technology Law*, 7 April 2022) <https://cyber.jotwell.com/the-disconnect-between-upstream-automation-and-legal-protection-against-automated-decision-making/> accessed 26 July 2025.
[75] EDPB/WP29 (n 53) 7; Elena Gil González and Paul De Hert, 'Understanding the Legal Provisions That Allow Processing and Profiling of Personal Data—an Analysis of GDPR Provisions and Principles' (2019) 19 ERA Forum 597, 608–609 <http://link.springer.com/10.1007/s12027-018-0546-z> accessed 23 June 2025.
[76] Council of Europe, *The protection of individuals with regard to automatic processing of personal data in the context of profiling: Recommendation CM/Rec(2010)13 and explanatory memorandum* (2010) para 95; Gil González and De Hert (n 75) 608–609.
[77] ibid.
[78] ibid.
[79] ibid 610-611.

3.1.1.4 Legal Effects and Similarly Significantly Effects

The scope of Article 22(1) GDPR extends to decisions that either produce legal effects or similarly significantly affect the data subject.[80] Legal effects typically arise where a decision affects an individual's legal rights or status, for example by granting or refusing social benefits, terminating a contract, or determining eligibility for a financial service.[81] The concept of "similarly significant effect" is broader, encompassing decisions that materially influence an individual's opportunities or obligations, such as access to housing, employment, or essential services.[82] Recital 71 GDPR illustrates this by mentioning the automatic refusal of an online credit application and e-recruiting practices.[83] The GDPR introduces the word "similarly", which is absent from Article 15 of the DPD, to ensure that the threshold for significance matches the seriousness of a decision producing a legal effect.[84]

Determining whether an effect is sufficiently significant requires a contextual, case-specific assessment.[85] Relevant factors include the type of personal data processed, the immediacy and duration of the consequences, whether the decision alters an individual's conduct or choices, and whether it results in a quantifiable loss or limits access to economic or social opportunities.[86] In the SyRI case in the Netherlands, for instance, a government-run algorithmic risk detection system generated profiles for fraud detection in disadvantaged neighbourhoods.[87] Although the system did not itself produce legal decisions, the court found that the algorithm's outputs had serious consequences for the private lives of those profiled and concluded that this violated the right to private life under Article 8 of the European Convention on Human Rights.[88]

Similarly, in the Ola case, the Amsterdam District Court held that automatically imposed fines and fare deductions on gig workers constituted a significant effect, as they altered the workers' contractual positions and affected their income.[89] In contrast, the same court found in two Uber-related cases that

[80] The GDPR does not define the term 'decision' for the purposes of Art 22(1). From the text of the provision, however, it is clear that the notion covers not only acts producing legal effects in relation to the individual concerned but also those that have a similarly significant impact on them; Case C-634/21 SCHUFA Holding [2023] ECLI:EU:C:2023:957, para 44.
[81] EDPB/WP29 (n 53) 21.
[82] ibid 21–22.
[83] Recital 71 GDPR.
[84] Article 15(1) DPD; EDPB/WP29 (n 53) 22.
[85] Sebastião Barros Vale and Gabriela Zanfir-Fortuna, 'Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities' (Future of Privacy Forum 2022) 35-37 <https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf> accessed 3 July 2025.
[86] ibid.
[87] Rechtbank Den Haag, Case C/09/550982 / HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (5 February 2020) para 6.59 <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1878>.
[88] ibid para 6.60.
[89] Rechtbank Amsterdam, Case C/13/689705 / HA RK 20-258, ECLI:NL:RBAMS:2021:1019 (11 March 2021) paras 4.51-4.52 <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2021:1019>.

temporary deactivations and ride-matching decisions, based on less intrusive data such as geolocation or traffic conditions, did not reach the threshold of Article 22 GDPR.[90] The effects in those cases were seen as reversible and not sufficiently substantial.

It is also important to consider how profiling and targeted communication can have disproportionate consequences for certain groups.[91] For example, advertising for high-interest financial products may appear innocuous at first glance, but when repeatedly directed at individuals in financial distress, it can exploit existing vulnerabilities and reinforce cycles of economic harm.[92] The threshold for a "similarly significant effect" therefore cannot be assessed solely in abstract terms. It requires sensitivity to context, including the structural position of the data subject and the broader social impact of the decision.

3.1.1.5 The Exceptions

While paragraph 1 establishes a general prohibition on decisions made solely through automated processing that produce legal or similarly significant effects, paragraph 2 provides exceptions that permit such processing under limited conditions.[93] These exceptions include situations where the decision is necessary for entering into or performing a contract,[94] where it is authorised by Union or Member State law,[95] or where it is based on the explicit consent of the data subject.[96] Each exception reflects a balance between enabling automation for legitimate purposes and protecting individual rights. However, these exceptions are not without controversy. Their application can undermine the protective intent of Article 22(1) GDPR, particularly when safeguards required under Article 22(3) are implemented in a formalistic or superficial manner. Recital 71 GDPR echoes the wording of Article 22(2) GDPR and states that such decision-making may be allowed under those conditions, but its phrasing is not legally definitive.[97] In fact, the expression “should be allowed” is often used in the recitals to illustrate possible applications rather than to mandate specific outcomes.[98] Some language versions of Recital 71 GDPR even frame the exceptions in terms of possibility or discretion, rather than

[90] Rechtbank Amsterdam, Case C/13/692003 / HA RK 20-302, ECLI:NL:RBAMS:2021:1018 (11 March 2021) para 4.26 <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2021:1018>.
Rechtbank Amsterdam, Case C/13/687315 / HA RK 20-207, ECLI:NL:RBAMS:2021:1020 (11 March 2021) paras 4.66-4.67 <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBAMS:2021:1020>.
[91] EDPB/WP29 (n 53) 22.
[92] ibid.
[93] Article 22(2) GDPR.
[94] Article 22(2)(a) GDPR.
[95] Article 22(2)(b) GDPR.
[96] Article 22(2)(c) GDPR.
[97] Christopher Kuner and others, ‘The EU General Data Protection Regulation: A Commentary/Update of Selected Articles’ [2021] SSRN Electronic Journal 97 <https://www.ssrn.com/abstract=3839645> accessed 27 April 2024.
[98] For example Recital 33 GDPR; Tosoni (n 54) 14.

necessity or entitlement.[99] The following subsections examine each exception in detail, analysing their scope, challenges, and implications for human oversight.

3.1.1.5.1 Explicit Consent

Explicit consent is one of the grounds upon which fully automated decision-making may lawfully occur. Consent under the GDPR must be freely given, specific, informed, and unambiguous, with explicit consent requiring a particularly clear and affirmative act by the data subject.[100] In theory, this mechanism is intended to give individuals control over the use of their personal data in high-risk automated contexts.[101] In practice, however, relying on consent as a safeguard raises significant concerns about whether such control is truly meaningful.

A key issue lies in the informational asymmetry between controllers and data subjects. For consent to be informed, controllers must provide clear and intelligible information about the existence of ADM, the logic involved, and the significance and envisaged consequences of the processing.[102] Yet, the opacity of many automated systems, especially those based on machine learning, makes it difficult for controllers to meet this standard and for data subjects to understand what they are agreeing to.[103] Even when information is provided, the complexity of the algorithms and their potential effects often exceed what most individuals can reasonably be expected to comprehend. This challenge is compounded in settings where individuals are pressured to consent due to financial incentives or unequal bargaining positions.

A recent decision by the Belgian DPA illustrates these concerns.[104] The GBA held that a data subject's consent to the processing of sensitive health data by an insurance broker was not valid because it was tied to receiving a discount on a mortgage interest rate.[105] The DPA found that the consent was not

[99] ibid 14.
[100] Article 4(11) GDPR; Recital 32 GDPR; EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' 6-7 <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en> accessed 28 June 2025.
[101] EDPB, 'Guidelines 05/2020 on Consent under Regulation 2016/679' para 13 <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en> accessed 28 June 2025.
[102] ibid para 62–63.
[103] Machine learning (ML) refers to a subset of artificial intelligence in which systems are trained to identify patterns in data and improve their performance without being explicitly programmed for every rule. ML-based systems are often opaque because they operate through complex, non-linear models (such as deep neural networks), making it difficult to trace how specific outputs are generated. This distinguishes them from rule-based systems, which follow clearly defined, human-coded instructions and are typically more transparent and predictable. See Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 2053951715622512 <https://doi.org/10.1177/2053951715622512> accessed 28 June 2025.
[104] Gegevensbeschermingsautoriteit (Belgium), Decision on the Merits No. 109/2024 (29 August 2024) <https://www.gegevensbeschermingsautoriteit.be/publications/beslissing-ten-gronde-nr.-109-2024.pdf>.
[105] ibid para 3.

freely given, as the data subject faced negative financial consequences if they refused, thereby violating the requirement of voluntariness under Articles 4(11) and 9(2)(a) GDPR.[106] This case highlights how, even when a formal consent mechanism is in place, structural dependencies can render it ineffective in practice.

The drafting history of European data protection legislation reflects these concerns. The European Commission resisted including consent as a justification for automated decisions during the negotiation of the DPD, arguing that power imbalances between data subjects and controllers could render consent ineffective as a safeguard.[107] Nonetheless, the GDPR introduced explicit consent as one of the exceptions.

Despite these difficulties, where explicit consent is relied upon, controllers are required to implement safeguards to protect the data subject's rights, freedoms, and legitimate interests.[108]

3.1.1.5.2 Contractual Necessity

The second exception allows for fully automated decision-making where it is necessary for concluding or carrying out a contract between the data subject and a data controller.[109] At first glance, this exception appears to provide a practical and convenient legal basis for the integration of automation into commercial services. However, its application is more limited than it may initially seem, because it has to be necessary. This exception applies only where the core subject matter of the contract cannot be fulfilled without the use of ADM.[110] The condition is not met simply because the controller finds automation useful, cost-effective, or efficient.[111] Rather, the necessity test is only satisfied when no viable alternative to automation exists for achieving the contract's purpose.[112]

This strict interpretation has been reinforced by regulators and courts assessing automated credit scoring practices.[113] The Finnish DPA, Data Protection Ombudsman, held that excluding applicants based solely on an upper age limit failed to meet the necessity requirement. The Ombudsman emphasised that a person's financial situation is more relevant to creditworthiness than demographic

[106] ibid para 5.
[107] Article 16 Amended proposal for a COUNCIL DIRECTIVE on the protection of individuals with regard to the processing of personal data and on the free movement of such data 1992.
[108] Article 22(3) GDPR; These safeguards are further discussed in Section 3.1.1.5.4.
[109] Article 22(2)(a) GDPR.
[110] EDPB/WP29 (n 53) 13.
[111] ibid 23.
[112] ibid.
[113] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 45.

factors such as age. For the exception to apply, the data used in the automated decision must be genuinely necessary for achieving the contract's objective.[114]

Other supervisory authorities have taken similar positions. The Icelandic DPA, Persónuvernd, confirmed that credit scoring constitutes profiling and that data subjects must be provided with a meaningful explanation of how such scores are generated.[115] This includes information on the logic and impact of the automated assessment. In a pre-GDPR ruling, the German Federal Court of Justice found that where a human decision-maker ultimately evaluates the result of a credit scoring system, the process does not fall within the scope of the GDPR's prohibition on fully automated decisions.[116]

These examples illustrate a recurring theme in the case law. The line between permitted and prohibited automation is often drawn based on how integral the automated system is to the contract's core purpose, and whether meaningful human involvement is retained. This scrutiny is particularly important in sectors like financial services, where automation can profoundly affect individuals' access to credit, housing, or other essential services. The rise of Smart Contracts and Agentic AI systems, which may autonomously execute transactions or enforce contractual obligations based on pre-set conditions, further heightens the need for meaningful human oversight.[117]

As with the consent exception, where the contractual necessity ground is invoked, the controller must implement safeguards to protect the data subject's rights.[118] However, if the necessity of the automation is not rigorously justified, the entire foundation for invoking the exception becomes questionable.

3.1.1.5.3 Authorisation by Law

The third exception permits fully automated decision-making when such processing is when it is explicitly permitted by EU or national law, provided that this legal basis also includes appropriate

[114] ibid 46.
[115] ibid.
[116] Bundesgerichtshof, Case VI ZR 156/13 (28 January 2014) para 34 <https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=66910&pos=0&anz=1>.
[117] Smart Contracts are self-executing digital agreements that automatically trigger actions once coded conditions are met, often deployed in blockchain environments. Agentic AI refers to systems with a degree of operational autonomy, capable of initiating actions and decisions without continuous human input. Their integration in financial infrastructures raises particular oversight concerns due to the potential for opaque and irrevocable transactions. See Bhabendu Kumar Mohanta, Soumyashree S Panda, and Debasish Jena, 'An Overview of Smart Contract and Use Cases in Blockchain Technology' (2025) <https://www.researchgate.net/publication/328581609_An_Overview_of_Smart_Contract_and_Use_Cases_in_Blockchain_Technology> accessed 22 July 2025; Izunna Okpala, Ashkan Golgoon and Arjun Ravi Kannan, 'Agentic AI Systems Applied to Tasks in Financial Services: Modeling and Model Risk Management Crews' (arXiv, 29 April 2025) <http://arxiv.org/abs/2502.05439> accessed 22 July 2025.
[118] Article 22(3) GDPR.

safeguards to protect the rights, freedoms, and legitimate interests of individuals.[119] This provision reflects the legislator's recognition that ADM may be justified in certain public interest contexts.[120]

Unlike the exceptions based on consent or contract, the legal authorisation ground is not left to the discretion of the controller but must be explicitly established in binding legislation. This requirement imposes a formal procedural threshold, ensuring that ADM processes with significant effects on individuals are not adopted in a legal vacuum. Such authorisation may be applied in areas like fraud prevention, tax surveillance, or efforts to ensure security and reliability, as concerns about efficiency and scalability have led governments to adopt algorithmic systems.[121]

However, the GDPR does not specify what constitutes "suitable safeguards," leaving their definition to the EDPB[122] and implementation largely to national legislatures.[123] This ambiguity doesn't facilitate the consistency and sufficiency of protections across different jurisdictions. The next section will explore the nature and scope of these safeguards in more detail.

3.1.1.5.4 Safeguards

Where ADM is permitted under one of the exceptions, Article 22(3) GDPR imposes a mandatory obligation on the controller to implement safeguards aimed at protecting the data subject's rights and legitimate interests.[124] These include, at a minimum, the right to obtain human intervention, to express one's point of view, and to contest the decision.[125] These safeguards are not optional. They form the minimum protection required when consequential decisions are made by automated means.[126]

Each of these rights aims to ensure that individuals retain some form of agency and procedural fairness in the face of automation. The right to human intervention requires that a person reviews the decision with the authority and ability to alter its outcome.[127] This review must involve a reassessment of all relevant data, including any new information submitted by the data subject.[128] For this process to be

[119] Article 22(2)(b) GDPR.
[120] The ratio legis is reflected in Recital 71 GDPR.
[121] Recital 71 GDPR.
[122] This responsibility of the EDPB can be found in Article 70(1)(f) GDPR.
[123] EDPB/WP29 (n 53) 27.
[124] ibid.
[125] Article 22(3) GDPR; Recital 71 GDPR.
[126] EDPB/WP29 (n 53) 27–28; Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398, 399 <https://www.sciencedirect.com/science/article/pii/S026736491730376X> accessed 1 July 2025.
[127] EDPB/WP29 (n 53) 27.
[128] ibid.

meaningful, the reviewer must be equipped to understand both the rationale behind the system's decision and the individual context, and must possess genuine authority to modify the outcome.[129]

The right to express one's point of view is intended to allow individuals to supplement the automated assessment with contextual information.[130] Yet this presupposes that the individual has access to the rationale behind the decision in the first place.[131]

The right to contest the decision completes the triad of safeguards. It is premised on access to an avenue for redress, either through internal complaint mechanisms or external oversight bodies. While the GDPR does not prescribe how this right must be operationalised, its effectiveness depends on the data subject having insight into how the decision was reached.

Recital 71 of the GDPR states that individuals should be able to obtain an explanation of the decision reached following automated processing. For a long time, the absence of this language in the operative provisions of Article 22 GDPR led to significant academic and regulatory debate over whether a "right to explanation" existed under the GDPR at all.[132] However, AG De La Tour of clarified in Dun & Bradstreet Austria that such a right is grounded in Article 15(1)(h) GDPR.[133] The Court held that data subjects must receive relevant, intelligible, and accessible information about the procedure and principles actually applied to produce a specific outcome.[134] This interpretation confirms that the right of access includes an individualised explanation, enabling the data subject to understand and meaningfully exercise their rights under Article 22(3) GDPR, including the ability to obtain human intervention, express their views, and contest the decision.

The EDPB has provided a non-exhaustive list of further safeguards that controllers should consider.[135] These include regular quality checks and algorithmic audits to detect discriminatory outcomes,

---

[129] ibid.
[130] 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, 20–21 <https://doi.org/10.1093/idpl/ipx005> accessed 28 April 2024.
[131] Veale and Edwards (n 126) 399–400.
[132] See Wachter, Mittelstadt and Floridi (n 125); Tiago Sérgio Cabral, 'AI and the Right to Explanation: Three Legal Bases under the GDPR', *Data Protection and Privacy* (2021) <https://www.researchgate.net/publication/349496743_AI_and_the_Right_to_Explanation_Three_Legal_Bases_under_the_GDPR> accessed 30 June 2025; Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond' (Social Science Research Network, 1 August 2017) <https://papers.ssrn.com/abstract=3124901> accessed 29 June 2025; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (Social Science Research Network, 13 November 2017) <https://papers.ssrn.com/abstract=3088976> accessed 24 June 2025.
[133] Case C-203/22 Dun and Bradstreet Austria [2024] ECLI:EU:C:2024:745, Opinion of AG De La Tour, para 67.
[134] Case C-203/22 Dun and Bradstreet Austria [2025] ECLI:EU:C:2025:117, para 77.
[135] EDPB/WP29 (n 53) 32.

mechanisms for independent third-party review, transparency tools that explain system logic in a comprehensible form, and accessible procedures to trigger appeals or corrective processes.[136]

3.1.1.6 Special categories

There is an additional layer of protection where ADM involves special categories of personal data.[137] In such cases, the prohibition of Article 22(1) GDPR applies with even greater force, and the conditions for lawfully engaging in such processing are stricter. Specifically, automated decisions that are based on sensitive data are only permitted when two cumulative requirements are fulfilled.[138]

First, the processing must fall under one of the exceptions listed in Article 22(2) GDPR. Second, the processing must also rely either on the data subject's explicit consent or on a legal basis grounded in substantial public interest established by Union or Member State law.[139]

This dual threshold reflects a heightened concern over the potential discriminatory effects of automated decisions involving sensitive data such as race, health status, political opinions, or religious beliefs.[140] Data controllers are expected to ensure that such decisions do not lead to discriminatory outcomes and must implement appropriate safeguards to prevent this risk.[141]

In practice, the notion of "explicit consent" in this context should be interpreted in line with the standards applicable under Article 22(2)(c) GDPR, including strict requirements for transparency and informed decision-making. Similarly, the "suitable and specific measures" required under Article 22(4) GDPR mirror the safeguards described in Article 22(3) GDPR, including the rights to human intervention, to express one's point of view, and to contest the decision.[142]

3.1.1.7 Protection of Children

Although Article 22 of the GDPR does not explicitly distinguish between adults and children, the specific vulnerabilities of children in the context of ADM and profiling require careful attention. The GDPR recognises that children may be less aware of the risks, consequences, and available safeguards related to data processing, and therefore deserve particular protection.[143] While not an outright prohibition, the GDPR strongly discourages subjecting children to decisions made solely by automated means when those decisions have legal or similarly significant effects.[144]

---

[136] ibid.
[137] As defined in Article 9(1) GDPR.
[138] Article 22(4) GDPR.
[139] Article 9(2)(a) and (g) GDPR.
[140] Article 9(1) GDPR.
[141] Recital 71 GDPR.
[142] EDPB/WP29 (n 53) 24.
[143] Recital 38 GDPR.
[144] Recital 71 GDPR.

Recent case law reflects this cautious approach. In France, an administrative court annulled a regional facial recognition pilot that used biometric scans to grant or deny school access, ruling that high school students could not provide valid consent due to an imbalance of power.[145] While the case was still pending, the CNIL, the French DPA, had already raised concerns, emphasising that the use of facial recognition in schools was particularly problematic given the sensitivity of biometric data and the vulnerability of children.[146] Similarly, a Swedish court upheld a fine against a school that used facial recognition to track student attendance, again finding that consent could not be considered freely given.[147] These cases illustrate how automated systems involving sensitive data can have serious consequences for children, and why strict safeguards are necessary.

Profiling in the context of children presents heightened risks, especially when used for purposes such as behavioural advertising, content personalisation, or encouraging in-app purchases.[148] In such cases, children may be particularly susceptible to manipulation, as they often lack the maturity to critically assess the intent and consequences of targeted messaging.[149] For instance, profiling in online gaming environments may identify children who are more likely to make purchases and target them with personalised offers to encourage spending.[150] Even if this does not initially appear to meet the threshold of legal or similarly significant effect, the cumulative influence on behaviour and decision-making may place the processing within the scope of Article 22 GDPR. To help protect children, the GDPR also supports developing sector-specific codes of conduct that define age-appropriate safeguards.[151]

3.1.2 Principles Supporting Human Oversight

Beyond the specific safeguards laid down in Article 22 GDPR, several foundational principles in data protection law provide structural support for human oversight in ADM. These principles, found primarily in Article 5 GDPR, help ensure that data-driven processes are fair, lawful, transparent, and accountable. While not exhaustive, the principles of fairness, transparency, lawfulness, and accountability are particularly relevant because they directly shape how ADM systems are designed,

[145] Tribunal Administratif de Marseille, La Quadrature du Net, No. 1901249 (27 Nov. 2020), para 12 <https://forum.technopolice.fr/assets/uploads/files/1582802422930- 1090394890_1901249.pdf>.
[146] Commission nationale de l'informatique et des libertés, Expérimentation de la reconnaissance faciale dans deux lycées : la CNIL précise sa position, 29 October, 2019. <https://www.cnil.fr/fr/experimentation-de-la-reconnaissance-faciale-dans-deux-lycees-la-cnil-precise-sa-position>.
[147] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 42.
[148] Article 29 Data Protection Working Party, 'Opinion 02/2013 on Apps on Smart Devices' 26 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf>.
[149] EDPB/WP29 (n 53) 29.
[150] ibid.
[151] Article 40(2)(g) GDPR.

deployed, and scrutinised.[152] This section examines how each of these principles contributes to embedding human judgment into ADM processes.

3.1.2.1 Fairness

The principle of fairness is a foundational safeguard in the GDPR, requiring that data processing practices, including those involving ADM, treat individuals justly and ethically.[153] In the context of ADM, this principle mandates that systems must not lead to unjust or discriminatory outcomes, particularly in sensitive areas such as education, welfare distribution, hiring, and credit scoring.[154] Fairness is closely linked to the prohibition of discrimination, the need for non-arbitrary decision-making, and the obligation to ensure that decisions align with the reasonable expectations of individuals.[155]

Several recent cases demonstrate how fairness is being operationalised in ADM contexts.[156] One of the clearest examples is the case involving the International Baccalaureate. Due to COVID-19, the IB cancelled exams and instead implemented an algorithm to assign grades based on historical school performance and "school context".[157] The Norwegian DPA, Datatilsynet, found this to be a breach of the fairness principle and the data by design obligation[158], noting that it disadvantaged students from historically underperforming schools, regardless of their actual academic capabilities.[159] The DPA concluded that this form of profiling led to potentially discriminatory outcomes and failed to reflect students' individual efforts and merits.[160]

The SyRI case also highlights how fairness can be undermined by discriminatory profiling. In addition to the court's finding of a violation of the right to private life under Article 8 ECHR, the Dutch Data Protection Authority separately concluded that the use of dual nationality as a risk factor in welfare assessments breached the GDPR's fairness principle.[161] The AP found that the processing was not only

[152] Other data protection principles, such as purpose limitation, data minimisation, storage limitation, and integrity and confidentiality, may indirectly support human oversight but are not analysed in this section due to their less direct operational relevance to oversight mechanisms.
[153] Article 5(1)(a) GDPR.
[154] Recital 71 GDPR.
[155] EDPB/WP29 (n 53) 9–10.
[156] Gianclaudio Malgieri and Frank Pasquale, 'Licensing High-Risk Artificial Intelligence: Toward Ex Ante Justification for a Disruptive Technology' (2024) 52 Computer Law & Security Review 105899, 9 <https://www.sciencedirect.com/science/article/pii/S0267364923001097> accessed 4 July 2025.
[157] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 13–14.
[158] Article 25 GDPR.
[159] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 14.
[160] ibid.
[161] ibid 36.

unnecessary but also resulted in discriminatory outcomes, particularly affecting marginalised communities. [162]

These cases illustrate that fairness is not limited to preventing explicit discrimination. It also encompasses broader concerns such as power imbalances, systemic opacity, and the divergence between individuals' expectations and the logic of automated systems. Although often conflated with transparency, fairness is emerging as a distinct and increasingly central principle in ADM enforcement. Traditionally underexplored in legal doctrine and case law, fairness is now being applied to address not only discriminatory outcomes but also subtler structural harms. The EDPB has stressed that fairness should be understood in conjunction with transparency and ethics, underscoring that ADM systems must be assessed not merely for technical accuracy but for their capacity to respect individual dignity and agency.[163] Still, the concept remains underdefined in the GDPR, which leaves substantial interpretive space for regulators and courts.[164] As jurisprudence develops, fairness may serve as a flexible normative lens to evaluate a wider range of harms arising from ADM, particularly where transparency alone proves insufficient.

3.1.2.2 Transparency

Transparency is a cornerstone of the GDPR and a key precondition for human oversight. Without sufficient insight into how decisions are made, individuals cannot meaningfully exercise their rights to intervene, express their views, or contest outcomes. In the context of ADM, the GDPR embeds transparency across several provisions.[165] Data subjects must be informed when automated processing is being used, including an explanation of the logic behind it, its importance, and the potential effects it may have on them.[166] They also have the right to access this information in cases where automated decision-making significantly affects them.[167]

The GDPR's transparency framework applies to all ADM, regardless of whether the decision falls under Article 22 GDPR. As the EDPB has emphasised, controllers must disclose the use of profiling or ADM in privacy notices, even if these practices do not result in legal or similarly significant effects.[168] This

[162] Autoriteit Persoonsgegevens, 'Boete Belastingdienst kinderopvangtoeslag' <https://www.autoriteitpersoonsgegevens.nl/documenten/boete-belastingdienst-kinderopvangtoeslag> accessed 4 July 2025.
[163] EDPB/WP29 (n 53) 9–10, 25–26.
[164] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 14.
[165] Article 13-14 GDPR.
[166] Article 13(2)(f), 14(2)(g) GDPR.
[167] Article 15 GDPR.
[168] Article 5(1)(a) GDPR; Recital 60 GDPR; EDPB/WP29 (n 53) 16–17.

includes meaningful information about how profiles are created, what categories of data are used, and how the profiling may affect the data subject.

Nonetheless, transparency remains difficult to operationalise in practice. One barrier is the inherent complexity of algorithmic systems, especially when decisions are based on statistical inferences or machine learning models that are not readily explainable.[169] As a result, the explainability of ADM has been the subject of extensive academic debate.[170] Another barrier stems from the invocation of trade secrets or intellectual property rights by controllers, which may be used to withhold crucial details about the underlying logic.[171] This tension has led to considerable regulatory and judicial attention.

One prominent example comes from Austria, in the Dun & Bradstreet Austria case, where a mobile phone provider refused a contract extension based on a low credit score.[172] The individual requested access to information about the logic used to generate that score, arguing that the explanation provided was vague and inadequate.[173] The Austrian DPA agreed, ordering the company to disclose meaningful information that would allow the individual to understand how their data was used in the automated assessment.[174] The matter was ultimately referred to the CJEU, which confirmed that controllers must provide relevant, intelligible, and accessible information about the actual procedure and principles applied to produce the outcome, rejecting overly general or opaque disclosures.[175] In the Austrian proceedings, the referring court had also sought clarification on whether transparency obligations require disclosure of input data, the parameters used for profiling, and the reasoning behind the result, as well as whether trade secrets can limit this right of access and whether comparative information, even in pseudonymised form, must be disclosed to assess the fairness and reliability of the decision.[176]

---

[169] Burrell (n 103) 4–5; Wachter, Mittelstadt and Floridi (n 130) 7.
[170] See Kevin Bauer and others, 'Expl(AI)n It to Me – Explainable AI and Information Systems Research' (2021) 63 Business & Information Systems Engineering 79 <https://doi.org/10.1007/s12599-021-00683-2> accessed 8 August 2025; Jennifer Cobbe, Michelle Seng Ah Lee and Jatinder Singh, 'Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021) <https://dl.acm.org/doi/10.1145/3442188.3445921> accessed 8 August 2025; Gianclaudio Malgieri, '"Just" Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation' (2021) 1 Law and Business 16 <https://sciendo.com/article/10.2478/law-2021-0003> accessed 8 August 2025.
[171] Wachter, Mittelstadt and Floridi (n 130) 21–27.
[172] Vienna Regional Administrative Court (Landesverwaltungsgericht Wien), decision of 23 October 2019, W256 2217011-1/11E, 4.
[173] ibid 3.
[174] ibid 1.
[175] Case C-203/22 Dun and Bradstreet Austria [2025] ECLI:EU:C:2025:117, para 55-57.
[176] Sebastião Barros Vale (n 4) 19.

In Sweden, the DPA fined Klarna Bank approximately 750,000 EUR for failing to meet transparency obligations in its use of ADM for credit applications and fraud detection.[177] Klarna's privacy notice mentioned the types of personal data processed but did not identify which factors were decisive or how the internal scoring model operated. The regulator found that this lack of detail prevented data subjects from understanding why certain decisions were made, in breach of Articles 13(2)(f) and 14(2)(g) GDPR. It held that meaningful information must include the categories of data used, how these influence the outcome, and whether certain circumstances always lead to rejection.[178]

These examples reveal a central challenge: while the GDPR demands meaningful information, it offers little guidance on what this entails in practice. The EDPB has filled part of this gap by recommending that controllers provide high-level but specific explanations. These should include the categories of data used, the reasoning behind their selection, how profiles are constructed, and how they influence decision-making.[179] Mathematical formulas or algorithmic code are not required, but abstract or generic disclosures are insufficient.

What emerges is a multi-layered understanding of transparency. At a minimum, data subjects must be made aware of ADM, informed of its purpose, and given a substantive account of how their data contributes to specific outcomes. This empowers oversight, facilitates accountability, and enhances trust in automated systems.[180] Transparency is therefore not only a procedural requirement but a substantive enabler of the other rights associated with human oversight. It ensures that the protections offered under Article 22(3) GDPR are not rendered illusory by informational opacity.

3.1.2.3 Lawfulness

The principle of lawfulness, requires that any processing of personal data must be grounded in one of the legal bases provided under Article 6 GDPR.[181] When special categories of data are involved, an additional basis under Article 9(2) GDPR is also required. The principle of lawfulness applies to all personal data processing activities, including profiling and ADM, regardless of whether the decision falls within the scope of Article 22 GDPR. It serves as a broader structural safeguard, ensuring that even partially automated decisions are grounded in a valid legal basis.

Recent jurisprudence demonstrates how the requirement for a valid legal basis plays a crucial role in regulating ADM. In the Dutch e-screener case, a psychological evaluation tool used in firearm licence

[177] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 21.
[178] IMY, *Decision in case DI-2019-4062 concerning Klarna Bank AB* (28 March 2022) https://www.imy.se/globalassets/dokument/beslut/klarna/beslut-tillsyn-klarna.pdf accessed 5 July 2025, 18.
[179] EDPB/WP29 (n 53) 31.
[180] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 18.
[181] Article 5(1)(a) GDPR.

applications was challenged for its reliance on automated assessments.[182] The court found that the processing did not fall within the scope of Article 22 GDPR because it was not solely automated.[183] Additionally, it points out that the ADM component was justified under Article 6(1)(c) GDPR, as the processing was deemed strictly necessary to comply with a legal obligation imposed on the controller under national firearms law.[184]

A similar approach was taken by the Slovak Constitutional Court, which assessed the use of automated tax fraud risk profiling through the e-kasa system.[185] Although the system's outputs were not themselves final legal decisions, the court held that the processing nonetheless required a clear legal basis under Article 6(1)(e) GDPR.[186] It stressed that such processing cannot be left to administrative discretion but must be explicitly authorized by legislation, especially when used by public authorities.[187]

The importance of lawfulness also featured in the litigation surrounding the Austrian AMAS system, which algorithmically assessed the employment prospects of job seekers. Initially, the Austrian DPA found that the system breached Articles 5(1)(a), 6(1), and 9(2) GDPR due to insufficient legal grounding.[188] However, on appeal, the Federal Administrative Court upheld the use of the system, accepting that the relevant provisions of the Austrian Labour Market Service Act provided an adequate legal basis under Article 6(1)(e) and 9(2)(g) GDPR.[189] This shows how disputes over the interpretation of lawfulness remain central to the governance of ADM, particularly in contexts involving social rights and vulnerable populations.

3.1.2.4 Accountability

Accountability plays an important role in ensuring that controllers proactively manage the risks of ADM systems. Controllers are under a duty to demonstrate that they comply with all data protection principles, including fairness, transparency, and lawfulness.[190] In the context of ADM, this obligation extends to implementing technical and organisational safeguards that enable meaningful human

[182] See Section 3.1.1.2.
[183] Rechtbank Den Haag, Case C/09/585239 / HA ZA 19-1221, ECLI:NL:RBDHA:2020:865 (11 February 2020) para 4.26 <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1013>.
[184] ibid para 4.25.
[185] Ústavný súd Slovenskej republiky, Case 492/2021 Z. z. (10 November 2021) para 147 <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2021/492/20211217>.
[186] ibid 141.
[187] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 15.
[188] ibid 16.
[189] ibid.
[190] Article 5(2) GDPR.

oversight. Recital 71 GDPR reinforces this, stating that appropriate mathematical or statistical procedures should be used to ensure fairness and prevent discriminatory effects.

One of the most important accountability tools in this context is the DPIA. A DPIA is mandatory for any processing that involves a systematic and extensive evaluation of personal aspects based on automated processing, including profiling, that produces legal or similarly significant effects.[191] It is clear from the wording of the provision that the obligation also applies to decision-making processes that are not solely automated, particularly where profiling plays a central role in producing legal or similarly significant effects.[192]

Recent enforcement actions have underlined the importance of DPIAs as a safeguard for high-risk ADM, just like the EDPB had clarified.[193] In its decision against Deliveroo, the Garante, the Italian DPA, found that the company had failed to conduct a DPIA before deploying the "Frank" algorithm, which assigned work shifts to riders based on availability and behaviour.[194] The Garante concluded that this omission violated Article 35 GDPR, particularly because the system involved large-scale processing, affected vulnerable individuals, and evaluated work performance in ways that significantly impacted riders' earning opportunities.[195]

Similarly, in the Slovak e-kasa case, the Constitutional Court held that the Tax Authority's use of a fraud detection algorithm lacked a proper impact assessment. The Court stressed that such automated profiling systems require a DPIA to assess their broader implications for human rights, even if they are not fully covered by Article 22 GDPR.[196] This includes evaluating the proportionality of the system, the legal basis relied upon, and the presence of sufficient oversight mechanisms.[197]

In addition to DPIAs, the GDPR also imposes further accountability obligations. One such requirement is the designation of a Data Protection Officer. A DPO must be appointed where the controller engages in regular and systematic monitoring of individuals on a large scale, particularly when profiling or automated decision-making is central to its activities.[198] This institutional safeguard complements the

[191] Article 35(3)(a) GDPR.
[192] EDPB/WP29 (n 53) 29–30.
[193] EDPB, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679' (2017) 8–11 <https://ec.europa.eu/newsroom/article29/items/611236> accessed 6 July 2025.
[194] Garante per la protezione dei dati personali, Ordinanza ingiunzione nei confronti di Deliveroo Italy S.r.l. (22 July 2021) [9685994] <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9685994>.
[195] ibid.
[196] Ústavný súd Slovenskej republiky, Case 492/2021 Z. z. (10 November 2021) para 98 <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2021/492/20211217>.
[197] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 26–27.
[198] Article 37(1)(b) GDPR; EDPB/WP29 (n 53) 30.

anticipatory function of DPIAs by ensuring ongoing internal oversight and expertise in data protection compliance throughout the lifecycle of ADM systems.

3.1.3 Rights Enabling Human Oversight

Human oversight in ADM is not only supported by structural principles but also by enforceable individual rights. These rights provide concrete mechanisms for data subjects to understand, challenge, and influence how automated decisions are made about them. While Article 22(3) GDPR establishes minimum safeguards where solely automated decisions are permitted, a broader set of rights reinforces oversight more generally. These include, but are not limited to, the rights of access, rectification, objection, and effective remedy. Each plays a distinct role in enabling agency, redress, and transparency in the face of automation. This section examines how these rights function in ADM contexts, both within and beyond the scope of Article 22 GDPR.

3.1.3.1 Right of access

The right of access is a foundational tool for enabling human oversight in ADM.[199] It allows data subjects to confirm whether their personal data is being processed, obtain a copy of that data, and receive essential information about how and why the processing occurs. This includes, in cases involving ADM or profiling, the right to obtain "meaningful information about the logic involved" in the processing, as well as "the significance and the envisaged consequences" of that processing for the individual.[200]

This provision is essential to ensuring that the safeguards listed under Article 22(3) GDPR, such as the right to obtain human intervention, express one's view, and contest a decision, can be exercised in a meaningful way. Without adequate access to information about how a decision was made, including the rationale and underlying data, individuals are left with procedural rights that are difficult or impossible to use effectively.

According to the EDPB, the information provided under Article 15(1)(h) GDPR must, where possible, go beyond general explanations and include specific insights into the reasoning behind a concrete decision concerning the data subject.[201] This includes an explanation of the logic involved, how personal data contributed to the outcome, and what the expected effects were. The EDPB also recalls in its transparency guidelines that this duty applies broadly, not only to ADM within the scope of Article

[199] Article 15 GDPR.
[200] Article 15(1)(h) GDPR.
[201] EDPB, 'Guidelines 01/2022 on Data Subject Rights - Right of Access' (2023) para 121 <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-012022-data-subject-rights-right-access_en> accessed 6 July 2025.

22 GDPR, but also to profiling in general.[202] It underlines the principle that individuals should never be taken by surprise by how their data is used.[203]

These principles were recently tested in a significant 2024 case before the Amsterdam District Court, which dealt with the platform X.[204] The court held that the company's shadow banning of a user, making their content invisible to others based on a content moderation algorithm, constituted ADM under Article 22(1) GDPR.[205] The user had not been informed of the decision, nor had they been provided meaningful information about the logic involved.[206] In response to the user's access request, the company offered only vague references to its privacy policy and invoked trade secret protections.[207] The court rejected these arguments, emphasizing that data subjects have a right to access their personal data and understand how decisions are made about them.[208] It ordered the platform to disclose information about the underlying logic, the specific personal data used, and any related internal scoring or labelling systems.[209] The ruling confirms that access rights apply even in technically complex or commercially sensitive ADM settings, and that companies cannot use generic policy references to avoid their legal duties.

This reasoning aligns with the approach adopted by the CJEU in the Dun & Bradstreet Austria case.[210] There, the CJEU clarified that Article 15(1)(h) GDPR entitles individuals to an individualised explanation of how an automated decision was reached. The AG had already underlined that such information must be intelligible, concrete, and relevant to the specific data subject, not merely abstract descriptions of general processing.[211]

Taken together, these developments reinforce that the right of access enables human oversight. It empowers individuals to understand and evaluate automated decisions that affect them and to seek redress where appropriate. Crucially, it acts as a gateway to other rights and safeguards by ensuring that the underlying decision-making process is transparent and contestable.

---

[202] EDPB/WP29, 'Article 29 Working Party Guidelines on Transparency under Regulation 2016/679' (2018) para 41 <https://www.edpb.europa.eu/system/files/2023-09/wp260rev01_en.pdf>.
[203] ibid.
[204] Rechtbank Amsterdam, Case 742407 / HA RK 23-366, ECLI:NL:RBAMS:2024:4019 (4 July 2024) paras 2.1-2.7 <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2024:4019>.
[205] ibid 4.26-4.27.
[206] ibid 4.27.
[207] ibid 4.15, 4.19.
[208] ibid 4.26-4.27.
[209] ibid 4.37.
[210] Case C-203/22 Dun and Bradstreet Austria [2025] ECLI:EU:C:2025:117.
[211] As discussed earlier in section 3.1.1.5.4.

3.1.3.2 Right to rectification

The right to rectification enables data subjects to correct inaccurate personal data and, where appropriate, to complete incomplete data.[212] In the context of ADM and profiling, this right plays a supporting role in human oversight by allowing individuals to challenge the factual basis of algorithmic decisions. This right is especially important in risk-scoring contexts where a minor factual error may significantly distort the output profile.

Article 16 GDPR applies to both the input data used to generate a profile and the output data, such as a categorisation or risk score, assigned to the data subject.[213] Profiling often involves predictions or generalisations, which can be based on limited, outdated, or inaccurate information. Rectification provides a channel for individuals to correct such errors, especially when the profile has significant consequences.[214]

Importantly, the right also includes the ability to supplement data with additional contextual information.[215] This is especially relevant where statistical profiles are technically correct but fail to capture individual circumstances. For example, a data subject placed in a health risk category might present more recent or specific evidence to challenge the assumptions behind that grouping. In this way, Article 16 GDPR offers a mechanism to intervene in the decision-making process and enhance the reliability of ADM outcomes.

While this right is less directly associated with procedural safeguards than access or objection, it contributes to a broader architecture of human oversight by enabling individuals to influence both the data and the reasoning that shape automated decisions.

3.1.3.3 *Right to object*

Data subjects have the right to object to the processing of their personal data, including profiling, when such processing is based on the controller's public interest task or legitimate interests.[216] This right is particularly relevant in ADM contexts that fall outside the scope of Article 22 GDPR, where fully automated decision-making is not at issue but where profiling still influences decision outcomes. Unlike the human intervention safeguard under Article 22(3) GDPR, which only applies in specific ADM cases, the right to object has broader applicability and can be exercised even in non-automated profiling contexts.

[212] Article 16 GDPR.
[213] EDPB/WP29 (n 202) 18.
[214] ibid 17.
[215] ibid 18.
[216] Article 21 GDPR.

Once the right is exercised, the controller must stop or refrain from processing unless it can demonstrate compelling legitimate grounds that override the rights and freedoms of the individual.[217] During this assessment, the data subject may also invoke the right to restriction, requiring the controller to limit processing until the objection has been resolved.[218] It may also be necessary for the controller to remove the relevant personal data.[219] This balancing test imposes a stricter threshold than the general legitimate interest assessment under Article 6(1)(f) GDPR. The controller must consider the necessity of profiling for its objective, assess the impact on the data subject, and document how these interests are weighed.

The right to object also applies unconditionally when personal data is processed for direct marketing purposes, including profiling for that purpose.[220] In such cases, the controller must immediately cease processing upon the data subject's request, without requiring justification or balancing of interests.[221] This right should be easy to exercise, available at any time, and offered free of charge.[222]

3.1.3.4 Right to an effective remedy

The right to an effective remedy ensures that the rights discussed in this section are not merely theoretical but enforceable. While this right does not itself create mechanisms for human oversight, it plays a crucial role when other protections fail to function as intended.[223]

Individuals must be able to bring complaints to a supervisory authority, challenge inaction or unsatisfactory decisions by that authority, and seek direct judicial redress against organisations responsible for unlawful processing.[224] These routes are especially relevant in the context of ADM, where data subjects may be denied access to meaningful information, left in the dark about the basis of decisions, or ignored when they attempt to contest outcomes.

Their relevance was illustrated in the Amsterdam District Court's 2024 ruling involving platform X.[225] After receiving no meaningful response to an access request concerning an automated restriction, the data subject successfully relied on Article 79 GDPR to obtain judicial relief. This case demonstrates how effective remedies function as a backstop, ensuring that the substantive rights tied to human oversight can be meaningfully exercised.

---

[217] Article 18(1)(d) GDPR.
[218] ibid.
[219] Article 17(1)(c) GDPR.
[220] Article 21(2); EDPB/WP29 (n 202) 19.
[221] ibid.
[222] Recital 70 GDPR.
[223] Articles 77-79 GDPR.
[224] ibid.
[225] See Section 3.1.3.1.

3.2 Key Limitations of Human Oversight Under the GDPR

While the GDPR establishes a legal framework for human oversight, its practical implementation reveals several structural, procedural, and psychological limitations. The following subsections identify the most significant shortcomings that hinder the effective use of oversight in ADM contexts. This analysis is informed by recent academic work, including Ben Green's critical examination of oversight policies in algorithmic governance, which highlights how formal safeguards may fail in practice.[226] Rather than offering an exhaustive account, the sections below focus on the core limitations most relevant to the GDPR's approach and its reliance on Article 22 GDPR as the central oversight mechanism.

3.2.1 The Structural Failure of Article 22

3.2.1.1 Narrow Scope and High Threshold

Article 22 GDPR is widely recognized as the backbone provision addressing ADM, yet its practical scope is considerably limited. The provision only applies when two cumulative thresholds are met: the decision must be "solely based on automated processing" and must produce "legal or similarly significant effects." As a result, many automated systems that materially affect individuals may fall outside its ambit, despite having potentially serious implications. Scholars and regulators alike have flagged these two conditions as creating systemic blind spots in the EU's governance of ADM.

Scholars have offered a sharp critique of the "solely automated" threshold, arguing that it provides superficial protections.[227] They note that high-stakes algorithmic systems used in sectors such as criminal justice and social welfare typically include some form of human involvement, whether nominal or symbolic, allowing them to escape the provision's scrutiny.[228] This legal formulation thus incentivizes minimal human presence to avoid classification as "solely" automated, encouraging what has been described as regulatory evasion through "rubber stamping."[229] This concern is reinforced by the observation that few automated systems involve decisions made entirely by algorithms, and that even fully automated systems in practice may fall outside the scope of the provision if presented as decision support tools rather than autonomous agents.[230]

The EDPB has sought to tighten the interpretive net, clarifying that the term "solely" does not exclude systems where human involvement is only tokenistic. They argue that human input must be

[226] Green (n 10).
[227] ibid 7–8; Veale and Edwards (n 126) 400–401.
[228] Veale and Edwards (n 126) 400.
[229] ibid; Green (n 10) 7.
[230] Veale and Edwards (n 126) 400.

meaningful, requiring the person involved to have both "authority and competence" to change the outcome. This view has been echoed in case law, such as the Amsterdam District Court rulings on Uber and e-screener, where nominal human involvement failed to prevent the courts from recognizing the decisions as automated.[231] These cases demonstrate that courts are increasingly assessing whether the human actor has the temporal, contextual, and substantive capacity to influence the decision, rather than merely appearing in the workflow.

Despite this evolving interpretation, the Article's formulation remains structurally flawed, since it still invites gaming of the system by incorporating superficial oversight. The critique is also of the post hoc nature of the right to human intervention under Article 22(3) GDPR, arguing that it places the burden of contestation on affected individuals after harm has already occurred.[232] Procedural obstacles, such as delays or lack of awareness, often render this safeguard ineffective in practice.

The second component limiting the scope of Article 22 GDPR is the requirement that the decision must produce "legal or similarly significant effects." Here too, ambiguity persists. The EDPB has clarified that "significant" effects extend beyond changes in legal status and may include decisions that materially alter a person's circumstances, opportunities, or obligations.[233] Examples include eligibility for credit, access to employment, or exclusion from services. This interpretation rightly extends the provision to many "nudging" and profiling-based practices that might otherwise escape oversight.[234]

Nonetheless, definitional uncertainties remain. The distinction between decisions that are legally binding and those that merely influence behaviour can be blurry, particularly in areas such as personalized advertising or dynamic pricing.[235] Moreover, there is an unresolved tension between the individualistic framing of data protection and the group-level harms that profiling can generate.[236] While the EDPB acknowledges that significant effects can be triggered by data drawn from others, such as postcode-based inferences in credit scoring, it stops short of offering clear standards for evaluating group-based impacts.

Despite these limitations, it is important to nuance the critique. Case law and regulatory guidance have helped clarify the scope of "solely" and "significant" effects. For instance, courts have confirmed that it is not the presence of a human in the process that matters, but whether that human plays a

[231] See Section 3.1.1.2.
[232] Green (n 10) 7.
[233] EDPB/WP29 (n 53) 21–22.
[234] Veale and Edwards (n 126) 401.
[235] For a helpful overview of these issues, see Latanya Sweeney, 'Discrimination in Online Ad Delivery' (2013) 56 Commun. ACM 44 <https://dl.acm.org/doi/10.1145/2447976.2447990> accessed 2 August 2025; Frederik Zuiderveen Borgesius and Joost Poort, 'Online Price Discrimination and EU Data Privacy Law' (2017) 40 Journal of Consumer Policy 347 <https://doi.org/10.1007/s10603-017-9354-z> accessed 2 August 2025.
[236] Green (n 10) 12.

substantive and autonomous role in shaping the decision. Supervisory authorities have also enforced GDPR violations through adjacent provisions, including transparency, lawfulness, and fairness, when Article 22 GDPR was deemed inapplicable.[237]

3.2.1.2 Lack of Proactive Design Obligations

While the GDPR allows for human intervention as a safeguard against solely automated decisions, it does not require that systems be developed in ways that structurally enable such intervention. There is no legal obligation for providers of ADM systems to embed human oversight capacities into the system architecture from the outset. Instead, oversight is treated primarily as an ex post safeguard, which activated after the decision has already been made and, potentially, after harm has occurred.

The only reference in the GDPR to system design requirements is the obligation to ensure data protection by design and by default.[238] However, this obligation is framed in broad and generic terms. It does not impose specific duties to build mechanisms that support meaningful human intervention or interpretability. This limitation is confirmed by the EDPB Guidelines on Data Protection by Design and by Default, which make no mention of human oversight in the context of ADM.[239] Despite offering detailed recommendations on embedding data protection principles into system architecture, the guidelines do not acknowledge the particular challenges posed by ADM or the need to facilitate intervention or control by human actors.

This absence reflects a deeper structural gap in the GDPR's approach to human oversight. The regulation does not incorporate a proactive governance model that ensures systems are designed to allow for scrutiny, contestation, or override. As discussed earlier, the Norwegian DPA's decision against the International Baccalaureate illustrates how the absence of design-stage obligations can result in opaque and discriminatory decision-making, where affected individuals are left unable to understand or contest the system's outputs.[240] In that case, data protection by design was invoked to support fairness and transparency, but only after the system had already been deployed and caused harm.[241] This reactive posture exemplifies the limits of GDPR's architecture when it comes to ensuring effective human oversight.

[237] As discussed in Section 3.1.2.
[238] Article 25 GDPR.
[239] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en> accessed 2 August 2025.
[240] As discussed in Section 3.1.2.1.
[241] Sebastião Barros Vale and Gabriela Zanfir-Fortuna (n 54) 14.

3.2.1.3 Absence of Organisational and Qualitative Oversight Requirements

The GDPR's approach to human oversight is further undermined by its failure to define the organisational and qualitative conditions under which oversight must occur. Even when oversight mechanisms are formally present, the regulation does not specify who is responsible for intervening, what skills they must possess, or how their involvement should be monitored and validated. As a result, any employee, regardless of training or impartiality, may serve as a human reviewer of algorithmic decisions. This omission allows oversight to be carried out by individuals whose qualifications, impartiality, or cognitive capacity may be inadequate for the task.

Moreover, human oversight is not embedded within the internal governance structures of data controllers. The GDPR treats oversight primarily as an individual right of the data subject, rather than as a systemic organisational responsibility.[242] There is no requirement to define roles, allocate decision-making authority, or support reviewers through procedures, guidance, or performance standards. Therefore, the GDPR offers no corresponding obligations to ensure that oversight is structurally feasible, adequately resourced, or independently exercised.

While the GDPR does not impose a general obligation to monitor the effectiveness of human oversight in ADM, internal actors such as Data Protection Officers may partially fill this gap. When appointed, DPOs can promote accountability by facilitating internal scrutiny and supporting cooperation with supervisory authorities.[243] They are also positioned to encourage the integration of quality controls into oversight processes. However, their effectiveness depends on their mandate, expertise, and independence.[244] In the absence of broader structural duties, such as logging, review protocols, or audits, there remains no systematic way to assess whether human intervention functions meaningfully in practice.

These governance gaps are particularly concerning in the context of hybrid decision-making systems, where human and automated components are combined.[245] Such systems may be presented as non-automated to avoid triggering Article 22 GDPR, even when the human role is minimal or constrained. The lack of transparency regarding how automation and human discretion interact, could make it

[242] Levitina (n 56) 127.
[243] Articles 37-39 GDPR.
[244] Levitina (n 56) 127–128.
[245] Koen Smit and Martijn Zoet, 'A Governance Framework for (Semi) Automated Decision-Making', *Proceedings of the Tenth International Conference on Information, Process, and Knowledge Management* (2018) 83 <https://www.researchgate.net/publication/325975696_A_Governance_Framework_for_Semi_Automated_Decision-Making> accessed 5 August 2025; Therese Enarsson, Lena Enqvist and Markus Naarttijärvi, 'Approaching the Human in the Loop – Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts' (2022) 31 Information & Communications Technology Law 123, 125–126 <https://doi.org/10.1080/13600834.2021.1958860> accessed 28 April 2024.

difficult for affected individuals to detect the presence of ADM, let alone challenge its outcomes. The GDPR does not regulate how ADM may influence or script the decision-making behaviour of human actors, nor does it mandate disclosure in cases where automation is only partially involved. This opacity also raises broader concerns about accountability, as it constrains the ability of individuals to contest decisions that may significantly affect their lives. Limited traceability and auditing capacities in ADM systems further impede the detection and correction of errors, while also complicating regulatory oversight and delaying the implementation of remedial measures. [246]

Finally, the regulation relies almost entirely on individual complaints to detect failures in human oversight. There are no mechanisms for collective representation, independent monitoring of oversight practices, or structural redress of systemic flaws.[247] The data subject must not only suspect that oversight was deficient, but must also navigate procedural obstacles to access information and seek remedies. This individualised enforcement model is poorly suited to identifying broader governance breakdowns or discriminatory patterns affecting groups.

3.2.2 The Misuse of Human Oversight

3.2.2.1 Oversight as a Formal Excuse

Human oversight is often presented as a mechanism that legitimises the deployment of controversial ADM systems rather than as a meaningful safeguard that improves outcomes. Some argue that human oversight policies tend to provide political and legal cover for flawed systems, allowing their adoption even in high-stakes public sector contexts where risks of bias or error are substantial.[248] Rather than prompting a reconsideration of the underlying algorithmic tools, oversight becomes a procedural checkbox that reinforces their continued use.

This tendency is particularly visible in how policymakers call for human overrides as evidence of effective oversight. Multiple guidelines suggest that if human decision-makers routinely accept algorithmic outputs without question, the system may be de facto solely automated and thus legally problematic.[249] Accordingly, policymakers encourage overrides to demonstrate that human actors retain discretion. At first glance, this emphasis seems to create a safeguard against flawed

[246] Levitina (n 56) 123.
[247] Stephan Dreyer and Wolfgang Schulz, 'The General Data Protection Regulationand Automated Decision-Making:Will It Deliver?' [2019] Bertelsmann Foundation; Levitina (n 56) 130–131.
[248] This reverts to the logic of the accountability-oriented goal, though in this instance it serves to excuse deficiencies in the system; see Fink (n 33) 7; Green (n 10) 10; Madeleine Clare Elish, 'Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (Pre-Print)' (Social Science Research Network, 1 March 2019) 20 <https://papers.ssrn.com/abstract=2757236> accessed 3 August 2025.
[249] EDPB/WP29 (n 53); 'Guidance on the AI Auditing Framework' (UK Information Commissioner's Office 2020).

algorithms.[250] Yet Green points out that human overrides do not address the root problems that make such overrides necessary in the first place.[251]

First, overrides are ineffective when the algorithm lacks trustworthiness. In such cases, the appropriate solution is not to rely on fallible human correction, but to improve or abandon the system altogether.[252] Second, even when human judgment is needed to consider factors excluded from the algorithm, people often fail to properly balance algorithmic advice with other inputs.[253] The tendency to over-rely on automated output, particularly when faced with opaque or technical systems, weakens the corrective function that human intervention is supposed to fulfil.

This logic is illustrated through judicial reasoning in State v. Loomis.[254] The case concerned the use of COMPAS, a proprietary risk assessment algorithm, in determining a defendant's sentence. Loomis challenged its use on due process grounds, arguing that neither he nor the court could meaningfully interrogate the system's reasoning due to its confidential nature.[255] The Wisconsin Supreme Court upheld the use of COMPAS in sentencing, despite recognizing that its proprietary nature prevents defendants and judges from understanding how the risk scores are generated.[256] Although this constraint seemed to preserve judicial discretion, it relied on the assumption that discretion could meaningfully counterbalance systemic bias or opacity. In practice, judges are often unaware of the influence of such tools or may exercise discretion in racially biased ways.[257] Thus, oversight becomes a way to validate flawed tools, rather than a standard for whether those tools should be used at all.

3.2.2.2 The Idealised View of Human Oversight

The GDPR implicitly idealizes human oversight as inherently beneficial by framing it primarily as an individual right to obtain human intervention to express his or her point of view.[258] It presupposes human oversight to be a legal and normative safeguard capable of counterbalancing the risks of ADM. Yet, scholars have criticized this idealization, arguing that it creates unrealistic expectations about

[250] Green (n 10) 9.
[251] ibid.
[252] ibid.
[253] Raja Parasuraman and Dietrich H Manzey, 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52 Human Factors 381 <https://doi.org/10.1177/0018720810376055> accessed 24 June 2025.
[254] *STATE v LOOMIS* (Supreme Court of Wisconsin).
[255] Trevor Hunt, 'How Artificial Intelligence Will Change Administrative Law: The Government of Canada's Directive on automated Decision-Making | DLA Piper' <https://www.dlapiper.com/en-us/insights/publications/2023/05/how-artificial-intelligence-will-change-administrative-law-in-canada> accessed 3 August 2025.
[256] *STATE v. LOOMIS* (n 254) [8,66].
[257] Green (n 10) 9–10.
[258] GDPR, art 22(3); Recital 71 GDPR.

human reviewers' ability and willingness to effectively scrutinize algorithmic outputs.[259] Mounting empirical, psychological, and structural evidence suggests that such oversight is often illusory.[260] Rather than guaranteeing fair outcomes, the provision of a reactive individual right may obscure deeper governance gaps and the necessity of systemic oversight.[261]

Ben Green's analysis highlights the fundamental lack of empirical support underpinning the GDPR's reliance on human oversight.[262] While the law presumes that human involvement can prevent or mitigate the harms of algorithmic systems, most studies indicate that individuals are ill-equipped to provide meaningful checks on automated outputs. The task of oversight often proves more difficult than the original decision-making itself, as people are expected to evaluate and, where necessary, override decisions made by complex systems that are presumed to outperform them in accuracy or efficiency. This dynamic has been described as the "ironies of automation".[263] It is the paradox that automating a process can make human supervision more demanding, not less.[264]

In practical terms, human overseers are required to exercise judgment over systems whose inner workings may be opaque, statistically complex, or contextually misaligned with the reality of their decisions. Contemporary algorithmic systems, such as complex machine-learning models used in predictive analytics, often function as "black boxes" whose inner workings are not entirely interpretable even to experts, let alone to laypersons or human reviewers without specialized knowledge.[265] The skillset and information asymmetry between overseers and the systems they monitor creates a profound structural mismatch.[266] Even when training or procedural constraints are introduced, such as mandatory override documentation or supervisor approval, they often fail to resolve the deeper problem: the impossibility of reliably evaluating a system that operates beyond

[259] Lazcoz and de Hert (n 39) 3–4; Andrew D Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (Social Science Research Network, 2 March 2018) 1090 <https://papers.ssrn.com/abstract=3126971> accessed 24 June 2025.
[260] Green (n 10); Ben Green and Yiling Chen, 'Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts' (2021) 5 Proc. ACM Hum.-Comput. Interact. 418:1 <https://dl.acm.org/doi/10.1145/3479562> accessed 3 August 2025; Andreas Holzinger, Kurt Zatloukal and Heimo Müller, 'Is Human Oversight to AI Systems Still Possible?' (2025) 85 New Biotechnology 59 <https://www.sciencedirect.com/science/article/pii/S1871678424005636> accessed 25 June 2025.
[261] Lazcoz and de Hert (n 39) 5.
[262] Green (n 10).
[263] Lisanne Bainbridge, 'Ironies of Automation' (1983) 19 Automatica 775 <https://www.sciencedirect.com/science/article/pii/0005109883900468> accessed 3 August 2025.
[264] ibid 775–776.
[265] Malgieri and Comandé (n 132) 2–3; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015) <https://www.jstor.org/stable/j.ctt13x0hch> accessed 7 August 2025; Philipp Schmidt, Felix Biessmann and Timm Teubner, 'Transparency and Trust in Artificial Intelligence Systems' (2020) 29 Journal of Decision Systems 260 <https://doi.org/10.1080/12460125.2020.1819094> accessed 8 August 2025.
[266] Green (n 10) 11.

one's cognitive capacity.[267] Efforts to improve oversight through design interventions like structured decision-support tools, preliminary judgment prompts, or improved training may yield incremental benefits but cannot fully overcome these embedded limitations.[268]

More fundamentally, the ideal of human oversight rests on an unstable normative foundation. The policy rationale for algorithmic decision-making is often grounded in its perceived objectivity and consistency, especially its ability to follow rules and apply them uniformly. Human oversight, by contrast, is celebrated for its discretion and contextual sensitivity. This sets up a contradiction: oversight is expected to simultaneously preserve the formal logic of the algorithm while also introducing flexibility.[269] As illustrated through the *Loomis* case, courts and policymakers often attempt to resolve this tension through circular logic, claiming that algorithms can guide decisions, but only if they don't actually influence outcomes. This creates a policy deadlock that undermines both accountability and effectiveness.

Critically, even if human oversight operated flawlessly, intervening when needed and supporting algorithmic outputs when appropriate, this idealised model could still perpetuate harm. It risks legitimising systems that are flawed not because of predictive inaccuracy, but because of their foundational goals or societal impact.[270] As Green argues, tools such as facial recognition, pretrial risk assessments, or welfare scoring systems may violate rights, reinforce discrimination, or expand surveillance regardless of how accurately they function.[271] Effective human oversight in these cases does not dismantle unjust architectures; it merely conceals them behind a façade of procedural legitimacy.

Central to the critique of human oversight's effectiveness is the concept of "automation bias," a cognitive bias wherein human reviewers disproportionately defer to automated recommendations, often neglecting critical assessment or independent judgment.[272] Automation bias arises because of psychological tendencies to trust technology excessively, exacerbated by inadequate training or a lack of technical expertise.[273] Automation bias transforms supposedly critical oversight processes into superficial exercises, creating the illusion of protection without meaningful intervention.[274] This

[267] ibid 11–12.
[268] See Linda J Skitka, Kathleen L Mosier and Mark Burdick, 'Does Automation Bias Decision-Making?' (1999) 51 International Journal of Human-Computer Studies 991 <https://linkinghub.elsevier.com/retrieve/pii/S1071581999902525> accessed 11 December 2024; Parasuraman and Manzey (n 253).
[269] Green (n 10) 12.
[270] ibid 10.
[271] ibid 12.
[272] Parasuraman and Manzey (n 253) 390–392.
[273] ibid.
[274] Lazcoz and de Hert (n 39) 2.

phenomenon undermines the rationale behind GDPR's oversight requirements, as evidenced starkly in real-world cases. For instance, In the Dutch child care benefits scandal, human reviewers consistently failed to critically interrogate the algorithmically generated fraud risk profiles, which disproportionately targeted vulnerable groups and minority communities.[275] Despite the formal presence of human reviewers, tens of thousands of families suffered severe consequences, including financial ruin, emotional trauma, and social exclusion.[276]

Beyond automation bias itself, reviewers may also exhibit other forms of flawed behaviour such as confirmation bias, their own bias, fatigue, or lack of engagement, especially when oversight is an ancillary task assigned without clear support or training.[277] These psychological and organisational factors further weaken the presumption that human intervention is inherently corrective.

In sum, the GDPR's approach to human oversight reflects a legal conception that does not fully account for the practical and psychological complexities of real-world implementation. It tends to overlook the structural mismatch between human cognitive capacities and the technical sophistication of algorithmic systems. Moreover, its normative justification for human oversight risks circular reasoning and may unintentionally legitimise systems that raise ethical concerns, rather than serving as a robust safeguard against them.

[275] 'Xenophobic Machines: Discrimination through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal' (*Amnesty International*, 25 October 2021) <https://www.amnesty.org/en/documents/eur35/4686/2021/en/> accessed 24 June 2025; 'Dutch Scandal Serves as a Warning for Europe over Risks of Using Algorithms' *POLITICO* (29 March 2022) <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/> accessed 24 June 2025.

[276] 'Xenophobic Machines: Discrimination through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal' (n 275) 13.

[277] Veronika Alexander, Collin Blinder and Paul J Zak, 'Why Trust an Algorithm? Performance, Cognition, and Neurophysiology' (2018) 89 Computers in Human Behavior 279, 287 <https://www.sciencedirect.com/science/article/pii/S0747563218303480> accessed 5 August 2025; Levitina (n 53) 125–126.

4 The AI Act's Contributions to Human Oversight

This chapter addresses the second sub-question of the thesis: What new provisions does the AI Act introduce specifically for human oversight? To answer this question, the chapter examines the relevant provisions of the AI Act, focusing on how human oversight is regulated and operationalised under this new legal framework. As the AI Act has only recently been adopted, there is no relevant case law to draw from. The analysis therefore relies on the legislative text itself and on academic commentary that has begun to assess its normative structure and regulatory implications. The chapter begins by explaining the AI Act's risk-based architecture, which determines the scope and application of oversight obligations. It then analyses what human oversight must achieve, when it must be exercised, and by whom. The chapter also identifies complementary provisions that support human oversight.

4.1 The Risk-Based Architecture of the AI Act

This section explains how the AI Act classifies AI systems by risk level, and how this classification determines the scope of oversight obligations under Article 14 AI Act. Understanding the structure and logic behind this risk-based model is essential to assessing how the AI Act contributes to human oversight and where limitations remain.

4.1.1 The Logic Behind the Risk-Based Approach

4.1.1.1 The Concept of Risk

Before turning to the tiered regulatory structure of the AI Act, one must first understand what the concept of 'risk' entails. Although the AI Act places risk at the centre of its regulatory design, it does not offer a very deep conceptual analysis of the term. It defines risk as "the combination of the probability of an occurrence of harm and the severity of that harm,"[278] echoing a general understanding of risk as the likelihood that a source of hazard will result in actual harm.[279] This includes both physical and non-physical forms of harm, such as infringements of fundamental rights.

This definition corresponds to the classical understanding of risk as a situation in which the possible outcomes are known and can, at least theoretically, be quantified. By contrast, 'uncertainty' refers to situations where the consequences of an action are foreseeable but cannot be measured with precision.[280] In some cases, the effects of a system remain entirely outside the current horizon of

[278] Article 3(2) AI Act.
[279] Martin Ebers, 'Truly Risk-Based Regulation of Artificial Intelligence - How to Implement the EU's AI Act' (Social Science Research Network, 19 June 2024) 7 <https://papers.ssrn.com/abstract=4870387> accessed 24 July 2025; Baruch Fischhoff, Stephen R Watson and Chris Hope, 'Defining Risk' (1984) 17 Policy Sciences 123 <https://doi.org/10.1007/BF00146924> accessed 25 July 2025.
[280] Frank Knight, *Risk, Uncertainty, and Profit* (1921) 20; Ebers (n 279) 7.

knowledge, which means that even the possibility of harm is not yet recognised.[281] These distinctions are especially relevant in the context of AI, where harm may emerge over time or only in particular use environments, and where statistical data is often unavailable or incomplete.

4.1.1.2 Key Actors

Before turning to the obligations surrounding high-risk AI systems and the requirements for human oversight, it is important to clarify the main institutional and operational actors defined by the AI Act. These actors form the backbone of the regulatory framework and determine how responsibilities are distributed across the lifecycle of an AI system.

4.1.1.2.1 Provider

Consider the example of a credit-scoring AI system used to evaluate whether individuals qualify for a loan. The provider in this case is the company that develops or commissions the development of the AI model and places it on the EU market under its own name or brand.[282] This might be a technology firm that builds a proprietary algorithm and licenses it to financial institutions. Even if the technical development is outsourced, the entity responsible for placing the system on the market remains the provider.[283]

4.1.1.2.2 Deployer

Once the AI system is sold to and integrated by a financial institution, that institution becomes the deployer. The bank that uses the credit-scoring tool to assess the eligibility of loan applicants does not develop the system itself, but uses it under its own authority to make decisions that affect customers.[284] The deployer is responsible for implementing the system responsibly, in accordance with both the provider's instructions and applicable legal obligations.[285]

4.1.1.2.3 Authorised Representative

Where the provider is established outside the European Union, such as a US-based credit scoring company, it must designate an authorised representative based in the EU.[286] This actor serves as the official point of contact for regulatory oversight and ensures that certain legal obligations are met on behalf of the provider.[287]

[281] Ebers (n 279) 7.
[282] Article 3(3) AI Act.
[283] ibid.
[284] Article 3(4) AI Act.
[285] Article 26 AI Act.
[286] Article 3(5) AI Act.
[287] Article 22 AI Act, concerning high-risk AI systems.

4.1.1.2.4 Market surveillance authorities

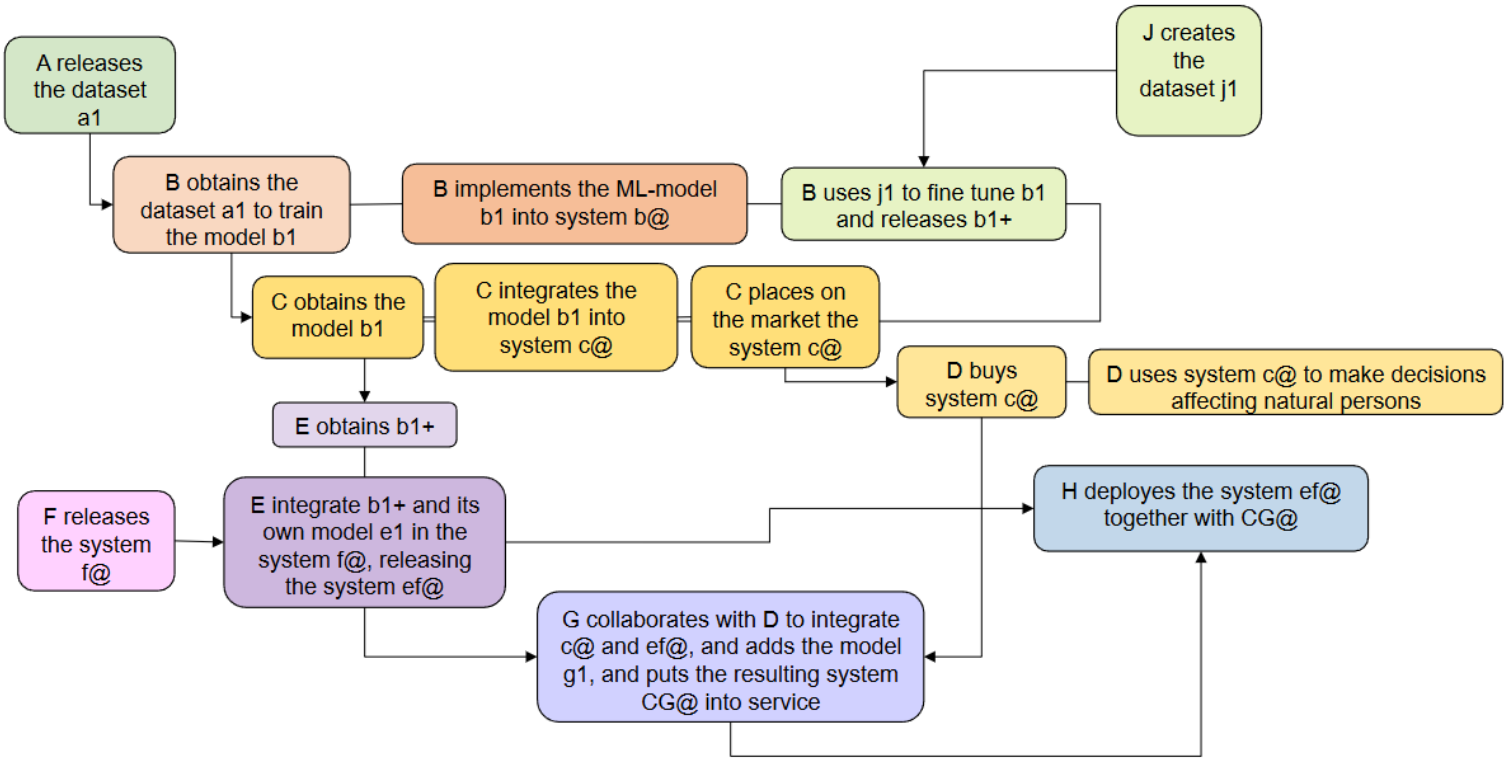
Market surveillance authorities are the national bodies appointed by each Member State to ensure that AI systems placed or used within their jurisdiction comply with the Regulation.[288] In this example, the authority could investigate whether the credit scoring tool is generating discriminatory outcomes or has been deployed without meeting transparency or human oversight requirements. In the context of AI systems used by EU institutions themselves, this role is carried out by the EDPS.[289]

4.1.1.2.5 Complexity of the AI Value Chain

Although the AI Act also defines other actors such as importers, distributors, and operators, their roles are less central to the oversight mechanisms addressed in this thesis and will not be explored in detail here.[290] While these actor categories may appear straightforward in isolation, they are part of a highly fragmented and multilayered AI value chain. A single AI system may involve multiple parties contributing to data input, model training, interface development, and system integration. AI is not a finished product or a static service, but a dynamic system delivered through multiple hands over time. It is a lifecycle that unfolds across diverse institutional settings, with varied effects on different individuals and groups.[291] In practice, distinguishing who qualifies as provider, deployer, or another actor is not always clear-cut. To exemplify this complexity, a schematic figure is included below to demonstrate how roles can overlap or shift across stages of the AI system lifecycle.[292]

[288] Article 3(26) AI Act.
[289] Article 70(9) AI Act.
[290] See Art 3(6)-(7)-(8) AI Act for the definitions of 'importer', 'distributor', and 'operator'.
[291] Lilian Edwards, 'Regulating AI in Europe: Four Problems and Four Solutions (Ada Lovelace Institute Expert Opinion)' (Social Science Research Network, 1 March 2022) 5 <https://papers.ssrn.com/abstract=5026691> accessed 21 July 2025.
[292] The figure was provided by Professor Gianmarco Gori during the course: *Legal Professionals in the Digital Age. (Regulatory Framework and Capita Selecta)* at the Vrije Universiteit Brussel.

SOFTWARE, DATA AND AI VALUE CHAIN



4.1.1.3 The Tiered Approach to AI Systems

Instead of relying primarily on liability mechanisms after harm occurs, the AI Act introduces a four-tiered risk classification system that reflects the European Union's commitment to proportionality and fundamental rights protection in technology governance.[293] Rather than imposing uniform rules, the Act adapts legal obligations to the level of risk posed by AI systems to human safety and fundamental rights. This stratified approach reflects a wider European regulatory logic that privileges preventive legal design over reactive enforcement, allowing regulatory intensity to vary based on the system's use context and societal implications. The classification scheme delineates four primary categories: prohibited AI, high-risk AI, limited-risk AI, and minimal-risk AI.[294] These categories are not arranged merely along a technological spectrum, but primarily in relation to the socio-legal context in which systems operate and the potential severity of their impact on individuals or society.

Prohibited AI systems are those deemed fundamentally incompatible with European values because they contradict the EU's core principles of respect for human dignity, freedom, equality, democracy, the rule of law, and fundamental rights enshrined in the CFR, including the rights to non-discrimination, data protection, privacy, and the rights of the child.[295] Banned uses include, among others, AI systems that manipulate human behaviour through subliminal techniques, exploit vulnerabilities based on age or disability, or engage in social scoring by public authorities.[296] These prohibitions, reflect the belief that some applications of AI cross an ethical threshold and cannot be justified even with safeguards.

High-risk AI systems form the core regulatory target of the Act.[297] These systems are typically deployed in areas such as biometric identification, critical infrastructure, education, employment, access to public services, law enforcement, migration and justice.[298] The logic underpinning their classification is rooted in their potential to produce significant harm, whether through direct physical danger or through subtle but profound effects on human rights, such as discrimination or the erosion of procedural fairness.[299] However, this rationale is not transparently articulated in the legislative documents. Neither the recitals of the AI Act nor any accompanying EU materials explain the criteria

[293] Article 5 TEU; Andrea Bertolini, 'Artificial Intelligence and Civil Liability' (Policy Department for Justice, Civil Liberties and Institutional Affairs 2025) 20 <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/776426/IUST_STU(2025)776426_EN.pdf>.
[294] Additionally, the regulation introduces a distinct category of 'general-purpose AI models', which will be examined in Section 4.1.2.3 below.
[295] Recital 28 AI Act.
[296] Article 5 AI Act.
[297] High-risk AI systems as defined in Art 6 AI Act; see further Section 4.1.2.1 below.
[298] Annex III AI Act.
[299] Enqvist (n 30) 515–517.

used to select these specific domains in Annex III.[300] The obligations imposed on high-risk systems are comprehensive, covering risk management, data quality, technical documentation, record keeping, transparency, human oversight, accuracy, cybersecurity and post-market monitoring.

Limited-risk systems are subject to lighter obligations, mainly transparency duties.[301] This category includes AI systems such as chatbots where the primary concern is user deception rather than systemic harm.[302] While the regulation does not explicitly refer to 'limited-risk' systems, it includes certain provisions that apply to them, such as the requirement to inform users that they are interacting with an AI system, thereby aiming to preserve informed consent and user autonomy.

Minimal or no risk systems comprise the majority of AI applications, including everyday tools like spam filters, video game AI, and weather prediction models.[303] These are effectively unregulated under the AI Act, though voluntary codes of conduct are encouraged. This exclusion from binding regulation is justified by their presumed low potential for harm and the desire to foster innovation where the risks are negligible.[304]

4.1.1.4 Why Human Oversight is Tied to High-Risk AI

The AI Act reserves human oversight obligations exclusively for high-risk systems. As codified in Article 14 AI Act, such obligations apply only to providers of high-risk AI.[305] This selective application reflects a regulatory judgment that oversight mechanisms are warranted only where the potential consequences of algorithmic malfunction, misuse, or bias are sufficiently serious to affect health, safety, or fundamental rights. Recital 52 of the AI Act makes it clear that high-risk systems are those likely to pose significant harm to these protected interests. In these settings, ADM without meaningful human involvement could result in unjust or irreversible outcomes. This limitation of binding oversight obligations to the high-risk category is also a product of regulatory pragmatism. Extending such requirements to all AI systems, regardless of their use case or level of autonomy, would risk overburdening developers and regulators alike, potentially stifling innovation in sectors where the societal impact of AI is limited. This would contradict the goal of the AI Act.[306] The AI Act adopts a

[300] Claudio Novelli and others, 'Taking AI Risks Seriously: A New Assessment Model for the AI Act' (2024) 39 AI & SOCIETY 2493, 13–14 <https://doi.org/10.1007/s00146-023-01723-z> accessed 24 July 2025.
[301] Article 50 AI Act.
[302] 'AI Act | Shaping Europe's Digital Future' (24 July 2025) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 24 July 2025.
[303] ibid.
[304] ibid.
[305] Article 6 in conjunction with Annexes I and III AI Act.
[306] As stated in Recital 2 AI Act, "… boosting innovation and employment and making the Union a leader in the uptake of trustworthy AI."

proportional approach, calibrating the level of regulatory intervention according to the degree of perceived risk and necessity.

It is worth noting, however, that this selective approach to oversight was not the only possibility considered during the legislative process.[307] The European Parliament's draft version of the AI Act proposed the insertion of a general principle of human oversight through Article 4a, which would have applied to all AI systems.[308] This provision, which ultimately did not survive the trilogue negotiations, would have required providers to make 'best efforts' to ensure human agency and oversight across the board.[309] However, these duties would have remained non-binding and largely aspirational. Ultimately, this amendment did not survive the interinstitutional negotiations. Practical concerns about enforceability, along with lobbying pressures from industry actors, likely contributed to its exclusion from the final text.

Nevertheless, the absence of binding oversight obligations outside the high-risk category does not preclude voluntary uptake of ethical practices. Providers and deployers remain free to integrate oversight features, assign responsible personnel, or design interfaces that allow for meaningful human control, even in cases where the law does not require it. In this regard, Recital 27 of the AI Act plays a significant signalling role. It recalls the seven non-binding ethical principles outlined in the 2019 Ethics Guidelines for Trustworthy AI, including human agency and oversight.[310] While these guidelines carry no legal force, they are explicitly encouraged as a basis for voluntary codes of conduct and best practices under the AI Act.

4.1.1.5 Preventive Orientation

A key aspect of the AI Act's regulatory logic is its preventive orientation. The AI Act moves away from ex post remedies, such as invoking human intervention after deployment.[311] Instead it requires

---

[307] Enqvist (n 30) 515–516.

[308] Amendments adopted by the European Parliament on 14 June 2023 on the proposal for the AI Act, amendment to add Article 4a,
available at <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html>
"General principles applicable to all AI systems
1. All operators falling under this Regulation shall make their best efforts to develop and use AI systems or foundation models in accordance with the following general principles establishing a high-level framework that promotes a coherent human-centric European approach to ethical and trustworthy Artificial Intelligence, which is fully in line with the Charter as well as the values on which the Union is founded:
a) 'human agency and oversight' means that AI systems shall be developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans;"

[309] Ebers (n 279) 14.

[310] High-Level Expert Group on AI (n 6).

[311] Such as provided in Article 22(3) GDPR.

oversight safeguards to be integrated proactively at the development phase.[312] This ex ante approach mandates that human oversight be embedded at the earliest stages of system design.[313]

This logic is reflected in Article 14 AI Act, which mandates that high-risk AI systems be designed to allow effective human oversight during use. The term 'effective' appears to align closely with the notion of 'meaningful', a standard that has been used to evaluate the quality of human involvement under the GDPR as well as in debates on autonomous weapons.[314] The growing emphasis on 'meaningful' human input reflects an awareness that not all forms of oversight are functionally adequate.[315] Superficial involvement, such as rubber-stamping decisions or being nominally in the loop, does not meet this threshold.[316] Influential bodies such as the Article 29 Working Party and the UK Information Commissioner's Office have stressed that oversight must involve genuine human judgment, discretion, and understanding of the system's functioning.[317] Meaningful oversight implies a capacity to critically engage with the system's output, not merely to observe or validate it without reflection.

While the operational responsibilities of providers and deployers will be examined in detail in Section 4.2, it is important to note here that oversight is not treated as an optional backstop, but as a structural feature of trustworthy AI.[318] This orientation reinforces the idea that meaningful human involvement cannot be an afterthought, but must be a core part of system architecture.[319] By placing oversight obligations at the design stage, the Act seeks to mitigate harm before it occurs.

4.1.2 Classification and Scope of High-Risk AI Systems

4.1.2.1 High-Risk AI Systems

The AI Act identifies high-risk AI systems through two distinct classification routes.[320] This structure allows the regulation to identify risk both through existing product safety regimes and through standalone use contexts where the deployment of AI poses particular threats to fundamental rights and safety.

---

[312] Lazcoz and de Hert (n 39) 6.
[313] Such as in Article 14(3) AI Act.
[314] Sarah Sterz and others, 'On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives', *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2024) 2 <https://dl.acm.org/doi/10.1145/3630106.3659051> accessed 22 December 2024; Fink (n 33) 3; Heather Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>.
[315] Green (n 10) 6–7.
[316] 'Guidance on the AI Auditing Framework' (n 249) 98.
[317] EDPB/WP29 (n 53); 'Guidance on the AI Auditing Framework' (n 249).
[318] High-Level Expert Group on AI (n 6) 15–16.
[319] ibid 21.
[320] Article 6(1) and 6(2) AI Act.

The first route concerns the embedded AI systems that function as safety components of products governed by existing Union harmonisation legislation.[321] These are listed in Annex I of the AI Act, and include instruments such as the Toys Safety Directive and Medical Devices Regulation.[322] When an AI system is either embedded in a product or is the product itself, and that product undergoes third-party conformity assessment, the AI system is also subject to the same scrutiny.[323] In this scenario, the risk classification stems not from the AI's functionality in the abstract, but from its integration into regulated physical goods. Examples include surgical robots, automated braking systems, or autonomous drones, where the AI plays a pivotal safety role and is regulated as part of a broader product safety framework.[324]

The second route, which has received greater public attention, is based on the standalone use of AI systems in sensitive application areas identified in Annex III. These include biometric identification and categorisation, education and vocational training, employment and worker management, access to essential private and public services, law enforcement, border control and migration, and administration of justice and democratic processes.[325] These categories do not presuppose physical danger, but are included because of the potential social, ethical, and fundamental rights impacts of using AI in such domains.[326]

4.1.2.2 Exceptions

Importantly, there are exceptions that allow certain AI systems listed in Annex III to escape high-risk classification, provided they do not pose a significant risk to health, safety, or fundamental rights.[327] Notably, this includes systems that detect patterns or deviations without replacing or materially influencing a prior human assessment, unless such assessment lacks proper human review.[328] In this way, meaningful human oversight serves not only as an internal safeguard within high-risk systems but also as a threshold criterion that determines whether the heightened regime of Article 14 AI Act applies at all.[329] However, this flexibility is not absolute. The Act explicitly states that any system performing profiling of natural persons is always to be treated as high-risk, regardless of its design or intended

[321] Article 6(1)(a) AI Act.
[322] Annex I (2) and (11) AI Act.
[323] Article 6(1)(b) AI Act.
[324] For example, on AI in robotics, see Martin Ebers, 'AI Robotics in Healthcare Between the EU Medical Device Regulation and the Artificial Intelligence Act' (2024) 11 Oslo Law Review 1 <https://www.scup.com/doi/10.18261/olr.11.1.2> accessed 25 July 2025.
[325] Annex III AI Act.
[326] Ebers (n 279) 15.
[327] Article 6(3) AI Act.
[328] Article 6(3)(c) AI Act.
[329] Article 6(3)(c) AI Act.

function.[330] Recital 53 of the AI Act reinforces this by linking the concept of profiling to Article 4(4) GDPR, thus importing a data protection rationale into the risk classification framework.

However, the exemption pathway provided by Article 6(3)(c) of the AI Act has raised concerns that it may function as a loophole. This provision excludes Annex III AI systems from high-risk classification if they are:

> *"intended to detect decision-making patterns or deviations from prior decision-making patterns and are not meant to replace or influence the previously completed human assessment, without proper human review."*[331]

This formulation introduces ambiguity. In practice, it could allow the system provider to define the AI's purpose as merely observational, not determinative, which could serve as a basis for avoiding the full set of obligations imposed on high-risk systems under Section 2 of the AI Act. By framing the AI's function as supportive rather than influential, providers might claim that human oversight neutralises the system's impact, thereby legitimising a downgrade in risk classification.[332]

At the same time, this potential loophole is constrained by the very phrasing of the provision. The exclusion applies only where the system is 'not meant' to replace or influence prior assessment, which appears to be an another intention-based criterion. Since this language centres on the declared purpose of the provider, it does not close the door to later accountability. Regulators and market surveillance authorities retain discretion to examine whether the system, in effect, does influence decision-making despite its stated intent. This scrutiny is grounded in procedural safeguards laid out in the AI Act: providers that consider their AI system exempt from the high-risk classification must document this assessment prior to making the system available on the market or putting it into service.[333] This assessment must be registered by the provider or authorised representative to the EU-wide database of high-risk AI systems, thereby creating a traceable compliance trail.[334] The database is publicly accessible, supporting transparency and enabling external accountability.[335] This does not apply to cases involving law enforcement, migration, asylum, or border control management, where confidentiality is maintained due to the sensitivity of the systems.[336]

[330] Article 6(3) AI Act.
[331] Article 6(3)(c) AI Act.
[332] Fink (n 33) 5.
[333] Article 6(4) AI Act.
[334] Article 49(2) AI Act.
[335] Article 71(2) and (4) in conjunction with Annex VIII, Section B(6) and (7) AI Act.
[336] Article 49(4) AI Act.

Besides, the AI Act does not operate in isolation. Even if a system avoids classification as high-risk under this provision, other legal regimes continue to apply. The GDPR remains applicable to any processing of personal data. National tort law, sectoral regulations, and principles of due diligence would also limit the leeway providers have in bypassing oversight requirements. In this sense, using Article 6(3)(c) AI Act as a tool to circumvent human intervention obligations would be not only legally questionable but strategically unwise.

In any case, the practical scope of Article 6(3)(c) AI Act appears limited at best. It only applies to AI systems listed in Annex III, a closed list focused on relatively specific AI systems. These systems already constitute a narrow subset of the overall AI market.[337] Additionally, it is difficult to identify real-world systems, that fall under Annex III, that would satisfy the dual requirement of detecting decision-making patterns or deviations, without replacing or influencing previously completed human assessment. The composite formulation creates a high bar for applicability. As a result, the actual number of systems that could credibly benefit from this exemption may be next to none. This casts doubt on the functional relevance of this provision, suggesting that it may be void rather than an exploitable loophole.

4.1.2.3 The Dynamic Nature of Annex III

Annex III of the AI Act provides a closed list of high-risk AI use cases, meaning that only those explicitly listed are classified as high-risk.[338] While this closed structure offers legal certainty, it also imposes rigidity. The exhaustive nature of the list limits the Act's responsiveness to emerging risks that may arise in unlisted domains. This rigidity risks leaving certain harmful or borderline applications outside the scope of mandatory oversight, even when they pose serious threats to individuals or society.[339]

To address this limitation, the European Commission can update Annex III through delegated acts.[340] While this mechanism theoretically allows the framework to adapt to evolving technologies and societal risks, the adoption of delegated acts is slow, politically sensitive, and evidence-dependent. This can hinder timely updates, particularly for general-purpose or adaptive AI.[341] Despite aiming to uphold the AI Act's risk-based logic, the procedure risks regulatory inertia and creates uncertainty due

[337] The Commission estimates that only approximately 5 to 15% of AI systems would fall under the high-risk category. Since Annex I systems are expected to account for the majority of these, the residual market share of Annex III systems is comparatively limited. See 'Impact Assessment of the Regulation on Artificial Intelligence' (European Commission 2021) 68–69.
[338] Article 6(2) AI Act.
[339] Ebers (n 279) 15.
[340] Article 7 AI Act.
[341] General-Purpose AI (GPAI) refers to AI systems that are designed for a wide range of tasks and can be used in various applications, not limited to a specific intended purpose. Examples include large language models like GPT-4, image generation models, or foundation models used across fields such as education, healthcare, and law; Article 4(63) AI Act; Adaptive AI refers to systems that can modify their behaviour after deployment without human intervention, often by learning from real-time data or user interactions.

to vague evidentiary standards. The Commission's gatekeeping role is therefore double-edged: it introduces adaptability in theory but may fail to deliver timely intervention in practice.

4.1.2.4 General-Purpose AI

The dual-path classification is not without challenges. In practice, many AI systems do not fit neatly into either Annex I or Annex III. This is particularly true for general-purpose AI systems, such as large language models, which are capable of performing multiple tasks across domains.[342] The AI Act's original framework struggled to accommodate such systems, prompting regulatory updates that introduced a separate compliance regime for general-purpose AI in Chapter V.

Even so, the boundary between general-purpose and high-risk remains porous. A general-purpose model may become high-risk once it is fine-tuned or deployed for a use case that falls under Annex III.[343] This dynamic classification process introduces significant uncertainty for providers, who may find that the regulatory status of their system depends not only on design but also on downstream use.[344]

4.2 Human Oversight Requirements for High-Risk AI Systems

Now that the scope of high-risk AI systems has been defined, we can turn to the core obligations surrounding human oversight. This section analyses how oversight must be implemented under the AI Act by addressing three interrelated questions: what functions oversight is meant to fulfil, when it must occur across the system lifecycle, and by whom it must be carried out. The dimensions of structure, timing, and responsibility form the basis of the following subchapters.

An instructive point of contrast with the GDPR emerges here. While Article 22 GDPR focuses on human oversight during decision-making, which is the use phase, the AI Act distributes oversight obligations across both development and use phases.[345] These phases are defined through the roles of provider and deployer, respectively.[346] Recital 73 further clarifies that the design of high-risk AI systems must already incorporate oversight mechanisms during development.

4.2.1 What Oversight Must achieve

4.2.1.1 Regulatory Purpose

To understand what human oversight entails under the AI Act, it is essential to examine the legally codified objectives set out in Article 14 AI Act. These provisions outline specific functional goals that

[342] Article 4(63) AI Act.
[343] Ebers (n 279) 16.
[344] Downstream use, as reflected in Article 3(68) of the AI Act, refers to how AI models are integrated, fine-tuned, or deployed into specific applications by entities (downstream providers) after their initial development; ibid.
[345] Lazcoz and de Hert (n 39) 6.
[346] Discussed in depth in Section 4.1.1.2

the oversight mechanism must enable. They require that natural persons assigned to oversee high-risk AI systems must be able to understand the system's relevant capabilities and limitations, monitor its functioning, interpret its outputs, and decide whether to act upon them.[347] Human oversight must be designed to prevent or reduce risks to health, safety, or fundamental rights that may arise during intended use or foreseeable misuse of high-risk AI systems.[348] In this sense, oversight is conceived primarily as a forward-looking harm-reduction mechanism and relates to minimising risks. This is very different from data protection law where human oversight is often associated with human dignity.[349]

These aims reflect a predominantly instrumental vision of oversight. The law frames oversight as a structured act of interaction between the system and the human operator, where the human assimilates and synthesizes outputs to detect potential harm.[350] Human oversight is expected to fulfil three key functions: identifying system errors or anomalies, preventing automation bias, and ensuring accurate interpretation of the system's outputs.[351] Rather than promoting open-ended human values, the AI Act focuses on ensuring reliable system performance and minimising harm in practice. This supports the Act's preventive regulatory logic, but also narrows the scope of oversight.

Fink's typology of human oversight identifies three distinct goals: output-oriented, process-oriented, and accountability-oriented oversight.[352] Human oversight in the AI Act primarily aligns with the first, focusing on mechanisms that prevent malfunction, misinterpretation, or over-reliance on AI outputs.[353] This reading is reinforced by the Commission's explanatory memorandum, which frames oversight as a means of reducing biased or erroneous decisions.[354] Yet elements of process-oriented oversight, such as autonomy, agency, and trust, are also reflected in surrounding policy documents and preparatory texts.[355] These highlight the importance of ensuring that users retain meaningful control over AI systems. By fostering user confidence, human oversight also contributes to broader

[347] Article 14(4) AI Act.
[348] Article 14(2) AI Act.
[349] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22 Computer Law Review International 97, 103 <http://arxiv.org/abs/2107.03721> accessed 24 December 2024.
[350] Enqvist (n 30) 520.
[351] Article 14(4)(a)-(c) AI Act; See for instance the insights from the Melanoma Patient Network Europe (MPNE), where patient advocacy experts emphasise the need for AI systems to produce results that are understandable to humans, especially in healthcare settings: 'MPNE Patient Consensus on Data and AI 2.0' (*Issuu*, 18 October 2024) <https://issuu.com/mpne/docs/mpne_consensus_2.0_long> accessed 29 July 2025.
[352] As discussed earlier in Section 2.3.
[353] As set out in Article 14(2) of the AI Act.
[354] European Commission, 'Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM(2021) 206 final, 2021/0106(COD).
[355] Riikka Koulu, 'Human Control over Automation : EU Policy and AI Ethics' (2020) 12 European journal of legal studies 30–31 <https://hdl.handle.net/1814/66992>; Fink (n 33) 9.

societal trust in AI systems.[356] Accountability-oriented goals, however, are absent from the text of Article 14 of the AI Act itself. While liability may still fall on human overseers in practice, the provision does not explicitly assign responsibility or establish structures of accountability.[357]

4.2.1.2 Structural Preconditions

While Article 14 of the AI Act defines the core tasks and objectives of human oversight, its meaningful execution requires more than a formal allocation of responsibility. Oversight must be supported by structural conditions that enable humans to understand, monitor, and, where necessary, intervene in AI system operation.[358] Among these conditions, access to clear, actionable information and retrospective traceability mechanisms are particularly important.[359] These provisions help ensure that oversight is not merely symbolic, but grounded in actual human capability. A more detailed discussion of these and other complementary provisions that support human oversight will follow in Section 4.3.

4.2.2 When Oversight Must Be Exercised

4.2.2.1 Temporal Dimensions

Determining when human oversight is to be performed, requires more than a focus on temporal immediacy. The question is not merely when in a chronological sense oversight should occur, but at which stage in the operation of a high-risk AI system a human should engage with its functioning.[360] As the sequence of system processes may include several interlinked components, oversight can take on different forms depending on whether it is directed at the system's internal mechanisms, its outputs, or its broader performance.[361]

The provision does not establish a fixed timeline or sequencing of oversight. While some obligations clearly relate to outputs, others apply more generally.[362] For example, the duty to enable human monitoring of the system's functioning and the ability to interrupt its operation are not bound to a specific moment.[363] This clause implies a form of oversight that spans the system's operational duration, rather than being limited to a discrete intervention point.

---

[356] Laux (n 26) 1–2.
[357] Fink (n 33) 9–10.
[358] Lazcoz and de Hert (n 39) 7.
[359] Article 13 and 14 AI Act.
[360] Enqvist (n 30) 525.
[361] ibid 526.
[362] 'Outputs' should not be understood narrowly as referring only to final decisions or recommendations. Article 3(1) AI Act provides a non-exhaustive list of examples, including content, predictions, recommendations, and decisions.
[363] This will be examined in greater depth in Section 4.2.2.2 below.

At the same time, the AI Act permits substantial discretion to providers regarding the operational design of oversight. This approach places considerable responsibility on system providers to determine when oversight should be exercised, both in terms of procedural integration and deployer instruction.[364] What qualifies as a system output, and thus the object of oversight, depends largely on provider-driven design decisions, including what data is presented to the deployer and when.[365] This design leeway also influences whether oversight is effectively performed at all. Although deployers technically retain the ability to intervene at any moment, they may not always have the contextual signals or clarity needed to know when such intervention is warranted.

4.2.2.2 Governance Models

The AI Act defines human oversight in functional rather than formal terms, allowing flexibility in how oversight is implemented across different stages of an AI system's lifecycle.[366] In practice, the timing of this obligation will depend on how oversight is structured throughout the lifecycle and operation of high-risk AI systems.[367] The oversight models introduced in Section 2.3.4 provide a conceptual framework for evaluating which arrangements are most consistent with the objectives and temporal structure.

Human-in-the-loop mechanisms, where a person must validate or guide every output before it takes effect, appear well aligned with the goal of enabling timely and informed intervention. Similarly, human-on-the-loop arrangements, which allow a person to monitor system operations and intervene when needed, may satisfy the oversight obligations under Article 14 AI Act. However, it remains unclear whether such real-time intervention capabilities are required in all high-risk cases.[368] As some scholars have noted, the text of Article 14(4) AI Act does not clarify whether the listed capabilities must be fulfilled cumulatively or whether they function as alternative requirements.[369] The only clear mandate for a human-in-the-loop model in the AI Act concerns remote biometric identification systems[370], where outputs must be verified and confirmed by at least two competent human reviewers.[371] This requirement does not apply in domains such as law enforcement or border control where Union or national law considers such verification disproportionate.[372]

[364] Enqvist (n 30) 526.
[365] ibid 527.
[366] Article 14(4) AI Act.
[367] Lazcoz and de Hert (n 39) 8.
[368] Sterz and others (n 314) 9.
[369] Fink (n 33) 3–4.
[370] Remote biometric identification systems do not include AI systems intended for the sole purpose to perform biometric verification, as clarified in Annex III, 1(a) AI Act.
[371] Article 14(5) AI Act.
[372] Art 14(5), second para, AI Act.

In practice, the answer likely lies somewhere in between. Article 14(4) AI Act specifies that human oversight must be ensured 'as appropriate and proportionate,' suggesting that not all systems must enable every listed capability in the same way.[373] Certain oversight features may be impractical or even hazardous depending on the context. For instance, requiring a stop button[374] for a surgical robot could endanger a patient mid-operation, while applying a hard stop to a self-driving car on a highway could result in severe accidents. In such scenarios, a proportionate response might involve a procedure that transitions the system into a safe fallback state, such as a controlled deceleration or the handover of control to a human operator under predefined conditions.

On the other hand, certain oversight elements, such as the ability to correctly interpret the system's output or to override its recommendations, seem indispensable for any meaningful human intervention.[375] Without these, the human operator would lack the tools necessary to exercise control or prevent harm effectively.[376] For this reason, assessing what constitutes appropriate and proportionate oversight cannot rely on a binary rule. A layered approach is more coherent: oversight requirements should first be considered at the level of system type. For example, one might assess what capabilities can reasonably be expected from all surgical robots or autonomous vehicles, and then refine those expectations on a case-by-case basis.[377] Within the same category of high-risk AI, the appropriateness of certain oversight functionalities may vary depending on the operational scenario. For instance, a full 'stop' mechanism might be unsafe during a critical phase of surgery where human intervention is not immediately feasible, while in another context, such as a less invasive procedure or at a different surgical stage, the same function could be safely triggered. This contextual calibration reflects the importance of tailoring oversight to specific system functionalities and deployment circumstances and indicates that flexibility is important in achieving the overarching objective of effective human involvement.

The human-out-of-the-loop model has already raised greater concerns as it can be seen as a second-step review.[378] Article 14 AI Act places emphasis on active monitoring and the ability to intervene before harm occurs. Oversight that is only possible after an output has taken effect fails to meet this anticipatory function. Similarly, human involvement limited to the development or testing phase, without any oversight capability during operation, does not fulfil the requirement of effective human

[373] Fink (n 33) 10.
[374] Pursuant Article 14(4)(e) AI Act.
[375] Article 14(4)(c)-(d) AI Act.
[376] Pursuant Article 14(1) AI Act.
[377] Sterz and others (n 314) 9.
[378] Lazcoz and de Hert (n 39) 8.

intervention. These forms of delayed or disconnected oversight do not align with the structure or intent of Article 14 of the AI Act.[379]

The human-back-in-control model, which involves an automatic transfer of control to a human operator under certain conditions, may offer a partial means of compliance. This presupposes that the system is designed to recognise predefined risk thresholds and to transfer control when they are met.[380] However, this model still requires that the human operator be sufficiently informed and capable of acting effectively once control is returned. Without such support, the risk remains that intervention comes too late or under conditions of uncertainty that impair its reliability.

Taken together, these models illustrate the range of potential implementation strategies, but they also highlight the limitations of relying on formalisms. The AI Act requires that oversight be effective, which implies not only the presence of a human overseer but the genuine capacity to understand and intervene in a timely and informed manner.[381] Rather than serving as fixed templates, the models discussed here should be used as analytical tools to evaluate oversight mechanisms in context. Determining whether a particular setup satisfies the legal standard must involve a case-by-case assessment, grounded in the specific risks, system functions, and operational conditions of each deployment.

4.2.3 Who Must Ensure Oversight

This section examines the actor-specific obligations for enabling human oversight under the AI Act. Unlike previous sections that considered oversight in terms of function or timing, this section focuses on the actors responsible for oversight: providers and deployers. The analysis draws on legal provisions[382], but also Fink's taxonomy of human oversight obligations in the EU AI Act and principles laid down by other scholars.[383]

4.2.3.1 Provider Responsibilities

Providers bear primary responsibility for designing AI systems with embedded oversight capabilities and supplying information that enables deployers to act responsibly.

4.2.3.1.1 Technical Enablers of Human Authority

While the functional importance of control mechanisms has already been discussed above, the focus here is on the provider's legal obligation to implement such mechanisms as part of the system's

[379] As discussed in Section 4.2.1.1 above.
[380] Lazcoz and de Hert (n 39) 7–8; Fink (n 33) 2–3.
[381] Article 14(1) AI Act.
[382] Mostly on Article 14 and 26 AI Act.
[383] Fink (n 33) 11; Sterz and others (n 314) 5; Enqvist (n 30) 528–531; Green (n 10).

technical design. This requires, as seen before, that natural persons tasked with oversight must be able to override or disregard the system's output, and to interrupt or stop the system through a mechanism that brings it to a safe state.[384] It is the provider who bears responsibility for embedding these capabilities directly into the system, but it is hard to understand how the provider can guarantee this.[385]

Fink observes that these requirements aim to ensure that human overseers retain authority over the AI system by making it possible to intervene in its operation or change its outcomes.[386] She notes that this includes both output-oriented and process-oriented controls, and that the provider must make such interventions technically possible within the design of the system.[387] However, the provision leaves providers with a degree of flexibility as to how these requirements are implemented, which raises questions about the limits of such discretion.[388]

From a broader perspective, the causal power can be identified as a key condition for effective oversight.[389] This refers to whether the human overseer has the real ability to alter the course of action taken by the system.[390] The inclusion of override and stop functions within the provider's responsibilities directly reflects this condition, as it ensures that human operators are not mere observers but retain a meaningful capacity to intervene.

The asymmetry between provider and deployer becomes particularly evident here. While the provider must make intervention technically feasible, it is the deployer who must assign oversight to individuals who possess the necessary authority and can exercise it in practice.[391] The distinction illustrates how shared oversight responsibilities operate along both design and deployment dimensions.

4.2.3.1.2 Design for Comprehension and Interpretation

A second responsibility of the provider is to design high-risk AI systems in a way that enables the human overseer to understand the system's functioning and to interpret its outputs appropriately.[392] This obligation requires the person assigned to oversight to understand the capacities and limitations of the system and to correctly interpret its outputs.[393] Although the functional implications of these

[384] Article 14(4)(d)-(e) AI Act
[385] Enqvist (n 30) 529.
[386] Fink (n 33) 11–12.
[387] ibid 11.
[388] ibid 10.
[389] Sterz and others (n 314) 5.
[390] ibid 6.
[391] Article 26(2) AI Act; Fink (n 33) 11.
[392] This relates to the transparency obligations that will be discussed in Section 4.3.1.1.
[393] Article 14(4)(a) and (c) AI Act.

requirements have already been discussed in previous sections, it is relevant here to emphasise the provider's role in enabling this comprehension through design.

Fink underlines that effective oversight requires more than the formal inclusion of such clauses. Providers must actually develop systems in a way that makes monitoring and interpretation feasible for the intended users.[394] This includes ensuring a degree of interpretability that allows users to understand the system's output well enough to act on it meaningfully and, where necessary, to justify or explain resulting decisions to affected individuals.[395] However, the law leaves open how such interpretability is to be achieved, especially in contexts where technical complexity makes system behaviour opaque. Yet, the ability to correctly interpret system output is not only a technical requirement but also a legal one, particularly where downstream actors bear explanatory or justificatory obligations.[396]

This concern is echoed by Sterz et al., who frame the issue in terms of epistemic access.[397] They argue that human oversight is ineffective unless the overseer can access and understand relevant system information.[398] This includes not only knowledge about what the system is doing but also why and how it reaches particular outputs.[399] Without such epistemic access, oversight risks becoming symbolic, since the overseer cannot meaningfully assess or challenge the system's behaviour.

4.2.3.2 Deployer Responsibilities

Deployers are responsible for assigning oversight to natural persons who can effectively exercise it. This implementation role is vital but underdeveloped in the AI Act's current formulation.

4.2.3.2.1 Equipping the Human Overseer

The AI Act establishes a duty on the deployer to assign human oversight to natural persons "who have the necessary competence, training and authority, and the necessary support."[400] As scholars have observed, it mirrors the functional structure of Article 14(4) AI Act and can be organised along three dimensions: authority, comprehension, and environment.[401]

[394] Fink (n 33) 12.
[395] This relates to the right to an explanation under Article 86 AI Act, which will be discussed in Section 4.3.1.3.
[396] As reflected in Article 14(4)(c) AI Act.
[397] Sterz and others (n 314) 5.
[398] ibid 6.
[399] ibid.
[400] Article 26(2) AI Act; Recital 91 AI Act.
[401] Fink (n 33) 10–11; Enqvist (n 30) 530.

Authority refers to the legal and organisational capacity of the person assigned to act on the system. The reference to 'authority' seems to entail that the designated individual is not merely present but able to override, disregard, or suspend the system's output where appropriate.

Comprehension concerns the requirement that oversight be assigned to persons with the "necessary competence and training."[402] Providers have to describe oversight measures clearly in the instructions for use.[403] The deployer's role is to ensure that human overseers have sufficient expertise and AI literacy to apply these instructions effectively, including interpreting the system's outputs and understanding its limitations.[404] AI literacy, as a concept, reflects a broader recognition that deployers must possess not only technical familiarity but also contextual understanding of how AI systems operate, what their limitations are, and how to exercise judgment when interacting with them.[405]

The reference to 'necessary support' appears to function as a baseline enabler. It ensures that the individual assigned to oversight has the practical means and equipment to exercise authority and apply the required competence. While concise, this provision captures the essential preconditions for making oversight operational, but it does not specify how these must be fulfilled or what standards apply.

4.2.3.2.2 Recognising Automation Bias

Fink associates the necessary support with the broader environment in which oversight is exercised, identifying it as one of three key components alongside authority and competence.[406] One important aspect of this environment is the risk of automation bias, which is explicitly addressed.[407]

This provision requires that human overseers remain aware of the tendency to automatically rely or over-rely on the outputs of high-risk AI systems.[408] Although this obligation is formally addressed to providers, it presupposes that overseers are trained and supported in a way that enables such awareness.[409] Training may help mitigate these behavioural risks, but evidence shows that it is often insufficient on its own.[410] The deployer must therefore ensure that oversight is assigned in conditions where the human operator is not structurally predisposed to defer to the system's outputs.

[402] Fink (n 33) 12.
[403] Article 13(3)(d) AI Act.
[404] Article 4 AI Act; Fink (n 33) 12.
[405] Article 3(56) AI Act; Recital 20 AI Act; Gültekin-Várkonyi (n 13).
[406] Fink (n 33) 11.
[407] Article 14(4)(b) AI Act.
[408] ibid.
[409] Which seems like an obligation of deployers under Article 26(2) AI Act.
[410] Green (n 10) 7–8.

The AI Act emphasises this concern in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons.[411] This formulation underscores that decision-support systems are especially vulnerable to symbolic human involvement.

4.3 Complementary Provisions Supporting Human Oversight

While Article 14 forms the core of the AI Act's human oversight obligations, it does not stand alone. This section examines adjacent provisions that structurally support or reinforce meaningful human oversight. Some of these obligations have been mentioned previously, but they are analysed here in closer detail to clarify their complementary role, without going into their full complexity, in order to maintain the thesis's focus on oversight.

4.3.1 Informational Provisions

4.3.1.1 Transparency

Transparency obligations under the AI Act form a central part of the infrastructure that enables human oversight.[412] In contrast to the GDPR, the AI Act is primarily concerned with systemic risk management and regulatory governance, rather than individual rights.[413] As such, it does not provide for a comprehensive set of individual transparency rights, nor does it offer the kind of granular guidance found in the GDPR on what constitutes effective transparency for data subjects.[414] The AI Act instead emphasises transparency between providers and deployers of high-risk AI systems, particularly through documentation and information-sharing aimed at supporting oversight and compliance.[415] Although it envisions harmonised transparency rules for certain AI systems, the emphasis is placed on institutional actors rather than on the individuals affected by system outputs.

Nonetheless, the AI Act recognises that transparency is also essential to human oversight. Providers of high-risk AI systems are required to ensure that their systems come with instructions for use that are clear, concise, and accessible to the individuals who will interact with them.[416] These instructions must include information on the system's capabilities and limitations, its intended purpose, the level of accuracy, and foreseeable circumstances that may lead to risks or errors.[417] The information must also

[411] Article 14(4)(b) AI Act.
[412] Recital 27 AI Act.
[413] Margot E Kaminski and Gianclaudio Malgieri, 'The Right to Explanation in the AI Act' (Social Science Research Network, 8 March 2025) 11 <https://papers.ssrn.com/abstract=5194301> accessed 7 July 2025.
[414] As illustrated by recent case law, such as Case C-203/22 Dun and Bradstreet Austria [2025] ECLI:EU:C:2025:117.
[415] Article 13 AI Act.
[416] Article 13(2) AI Act; Sebastian Felix Schwemer, Letizia Tomada and Tommaso Pasini, 'Legal AI Systems in the EU's Proposed Artificial Intelligence Act' (Social Science Research Network, 21 June 2021) 6 <https://papers.ssrn.com/abstract=3871099> accessed 1 August 2025.
[417] Article 14(4) AI Act.

be relevant and accessible to the deployer, enabling them to interpret system outputs and use them appropriately.[418]

Human overseers are therefore expected to understand the system's functions and limits, monitor its behaviour, and interpret its outputs accurately.[419] These tasks rely on the quality and usability of the information provided.[420] At the same time, there is a burden on deployers to operate high-risk AI systems in accordance with the instructions for use.[421] The effectiveness of oversight therefore hinges on a shared responsibility between providers and deployers. Without a guarantee that the transparency measures are tailored to the deployer's role, expertise, or working environment, the responsibility assigned to deployers may become difficult to fulfil in practice.

While the transparency requirements appear extensive, the law offers little guidance on their practical implementation.[422] It does not specify how this information should be structured, presented, or validated for usability in actual deployment contexts. Nor is there any requirement to assess whether deployers actually understand the documentation. This results in a gap between disclosure and actionable understanding.

The emphasis on upstream documentation for compliance assessments, coupled with the absence of formalised usability testing, interface design standards, or feedback mechanisms, leaves providers with considerable discretion to define what counts as adequate transparency.[423] Without external validation, there is a risk that providers meet formal requirements without enabling meaningful human oversight.

4.3.1.2 Logging and Traceability

Logging obligations under the AI Act complement transparency by providing an ongoing, verifiable record of system behaviour.[424] While transparency focuses on ex ante access to operational information, record-keeping enables retrospective scrutiny of how high-risk AI systems were used, by whom, and under what conditions. Comprehensible records of a high-risk AI system's development and performance are essential to verify compliance, ensure traceability, and support both oversight

[418] Article 13(2) AI Act.
[419] Article 14(4) AI Act.
[420] Article 13 AI Act.
[421] Article 26(1) AI Act.
[422] Article 13(3) AI Act.
[423] Madalina Busuioc, Deirdre Curtin and Marco Almada, 'Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act' (2023) 2 European Law Open 79, 97–98 <https://www.cambridge.org/core/journals/european-law-open/article/reclaiming-transparency-contesting-the-logics-of-secrecy-within-the-ai-act/01B90DB4D042204EED7C4EEF6EEBE7EA> accessed 29 July 2025.
[424] Article 12(2) AI Act.

and post-market monitoring.[425] These logs must contain sufficient detail to trace system performance, identify errors, and evaluate the context in which outputs were generated. Although broadly framed, the obligation is particularly pertinent where human oversight is intended to be reviewable after the fact.

A more explicit link to human oversight is found in the requirement to maintain records identifying the natural persons responsible for verifying the results of remote biometric identification systems.[426] This refers to the requirement that an identification output must be independently confirmed by at least two human reviewers before any action can be taken based on it.[427] Here, logging serves not only a technical function but also a legal one. It provides evidence that oversight was exercised by competent, authorised individuals, as required under the AI Act.[428]

4.3.1.3 Right to Explanation

The AI Act establishes a right to receive a meaningful explanation about the role of an AI system in decision-making.[429] This complements human oversight by enabling individuals to understand how an AI system contributed to a decision and by whom responsibility was exercised. Importantly, unlike Article 22 of the GDPR, which applies only to fully automated decisions, the AI Act grants a right to explanation even when a human is involved in the decision-making process.[430] This broader scope reinforces transparency and contestability within human-in-the-loop systems, rather than weakening them.[431]

The relationship between Article 86 AI Act and the GDPR creates a dynamic interaction: "a hydraulic effect," as Kaminski and Malgieri call it.[432] The stronger the right to explanation under the GDPR, the narrower the added value of Article 86 AI Act. Conversely, if the GDPR right is interpreted narrowly, Article 86 AI Act assumes a broader remedial role. They also argue that Article 86(2) AI Act must not be misread as an automatic carve-out in favour of existing data protection law.[433] It should be interpreted narrowly to avoid undermining transparency in higher-risk AI systems while paradoxically strengthening it in lower-risk settings.[434]

---

[425] Recital 71 AI Act.
[426] Article 12(3)(d) AI Act.
[427] Article 14(5) AI Act, see also Section 4.2.2.2 above for further discussion.
[428] This clearly aligns with the requirement imposed on human overseers under Art 26(2) AI Act, as discussed in Section 4.2.3.2.2.
[429] Article 86(1) AI Act.
[430] Kaminski and Malgieri (n 413) 5–6.
[431] ibid 5.
[432] ibid 25–26.
[433] ibid 24–25.
[434] ibid 25.

4.3.2 Fundamental Rights Impact Assessment

4.3.2.1 Legal requirements

Deployers of certain high-risk AI systems must conduct a Fundamental Rights Impact Assessment prior to deployment.[435] This obligation applies only in specific situations, namely when the deployer is a body governed by public law or a private entity providing public services, and when AI systems are used to evaluate the creditworthiness of natural persons or for credit scoring. It also applies to AI systems used for risk assessment and pricing in life and health insurance, excluding those used solely for detecting financial fraud.[436] This obligation establishes a structured process for assessing whether an AI system may pose risks to fundamental rights, democracy, or the rule of law.[437] Unlike the GDPR's DPIA, which focuses on data protection, the FRIA is expressly concerned with the broader spectrum of fundamental rights, such as non-discrimination, access to justice, and freedom of expression. The FRIA must assess not only the purpose and context of use, but also whether the deployment is necessary and proportionate in light of its impact on individuals and society.

This obligation represents a clear shift toward embedding rights-based concerns at the heart of AI governance. The FRIA must contain information about the system's intended purpose, the context of use, the categories of persons affected, the specific rights at stake, and the measures adopted to address those risks.[438] It must be conducted before the system is put into use, and must be documented and kept available for supervisory authorities.[439]

4.3.2.2 Interplay with DPIA

Although the FRIA and the DPIA share structural similarities, they differ in scope, legal basis, and institutional reach. The DPIA under applies to any data controller whose processing is likely to result in a high risk to the rights and freedoms of natural persons.[440] In contrast, the FRIA is more limited.

This results in a clear asymmetry between the two instruments and may lead to an accountability gap for private sector actors deploying high-risk AI systems. Moreover, the AI Act refers to the DPIA explicitly, requiring deployers to indicate whether a DPIA has been carried out. In this regard, if a DPIA has already been conducted under Article 35 of the GDPR or Article 27 of the LED, the FRIA is required

[435] Article 27(1) AI Act.
[436] ibid.
[437] Recital 96 AI Act.
[438] Article 27 AI Act ; Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 Computer Law & Security Review 106020, 9–10 <http://arxiv.org/abs/2411.15149> accessed 1 August 2025.
[439] ibid 4–5, 9.
[440] Article 35(1) GDPR.

to complement that assessment rather than replace it.[441] Conflicts may also emerge between Conformity Assessment standards and the FRIA methodology, since Article 27 AI Act permits deployers to determine their own methodological approach, which is similar to the flexibility granted in data protection through the DPIA.[442]

[441] Article 27(4) AI Act.
[442] Mantelero (n 438) 6.

5 Evaluating the Shift in Human Oversight

This chapter addresses the main research question of the thesis: To what extent does the AI Act address the limitations identified in the GDPR's approach to human oversight, and what gaps remain? While earlier chapters focused on the legal design and operational rules of each instrument in isolation, this chapter takes a broader perspective to assess whether the AI Act provides a meaningful response to the oversight limitations identified under the GDPR. Building on the limitations outlined in Section 3.2, and the structure and obligations analysed in Chapter 4, the following sections evaluate how the two frameworks interact, where the AI Act introduces new safeguards, and where gaps or dependencies on the GDPR persist.

5.1 Context

5.1.1 Overview and Recap

As analysed in Section 3.2, the GDPR's approach to human oversight in ADM suffers from several critical limitations. Although Article 22 GDPR establishes a formal right to obtain human intervention, its effectiveness is constrained by structural, organisational, and psychological shortcomings. These issues are particularly pressing in high-stakes decision-making environments, where oversight must operate as more than a procedural safeguard.

The first limitation lies in the narrow legal scope of Article 22 GDPR. Its dual threshold requires decisions to be both solely automated and to produce legal or similarly significant effects. This excludes many impactful systems from scrutiny. This also allows for regulatory evasion through minimal human involvement and leaves oversight reactive, placing the burden of contestation on individuals after harm has occurred.

Second, the GDPR fails to require that human oversight be embedded into the design and governance of ADM systems. There are no obligations to ensure interpretability, allocate responsibilities, or monitor whether oversight is actually used or effective. Reviewers may lack the training, authority, or institutional support to carry out their role meaningfully. In hybrid systems, where automation and human discretion intersect, the absence of transparency further obscures accountability.

Third, the regulation rests on an idealised vision of human oversight. It assumes that human judgment is inherently corrective, despite mounting evidence that oversight may be symbolic, ineffective, or distorted by any type of bias. In some cases, oversight serves more to legitimise flawed systems than to mitigate their risks.

Taken together, these limitations point to a broader regulatory failure: the GDPR treats oversight as a static legal entitlement rather than a dynamic and embedded governance function. The following sections, 5.2 and 5.3, assess to what extent the AI Act provides a more robust framework capable of addressing these deficits and enabling meaningful human control.

5.1.2 Distinguishing Automated Decision-Making and High-Risk AI

5.1.2.1 Conceptual and Legal Divergences

The concepts of ADM under Article 22 GDPR and High-Risk AI under Article 6 and Annex III of the AI Act are distinct, though partially overlapping. While both frameworks derive from the TFEU and aim to protect individuals from the risks of data processing and algorithmic decision-making, they operate based on fundamentally different logics and trigger mechanisms.[443]

As previously discussed, the GDPR defines ADM narrowly. This creates a high threshold for applicability and relies on an effect-based trigger, where the nature and consequences of the decision determine whether Article 22 GDPR is engaged. By contrast, the AI Act classifies certain systems as high-risk based on their intended purpose and area of use, rather than the decisional outcome they produce.[444]

Crucially, the AI Act is "generally agnostic regarding the decision-making scheme vis-à-vis individuals when the deployment of an AI system is at stake."[445] It covers software capable of predictions and recommendations, which courts and data protection authorities have not necessarily considered fully automated decisions under the GDPR.[446] Furthermore, the rules governing high-risk AI systems under Sections 2 and 3 apply even when a human ultimately makes the final decision on behalf of the user, based on suggestions, signals, prompts, or recommendations generated by the AI system.[447]

This means that the scope of the AI Act may exceed that of Article 22 GDPR. The AI Act also does not necessarily ensure that its oversight mechanisms displace those of the GDPR. Incorporating human oversight in a high-risk AI system does not, for example, remove the possibility that Article 22 GDPR may still apply. As demonstrated by the Portuguese proctoring case, and the Dutch e-screener case, the presence of a human in the loop with the authority and ability to make final decisions does not

[443] Article 16 TFEU.
[444] Article 6 AI Act.
[445] Sebastião Barros Vale, 'GDPR and the AI Act Interplay: Lessons from FPF's ADM Case-Law Report' (*Future of Privacy Forum*, 2022) <https://fpf.org/blog/gdpr-and-the-ai-act-interplay-lessons-from-fpfs-adm-case-law-report/> accessed 5 August 2025.
[446] Article 3(1) AI Act.
[447] Sebastião Barros Vale (n 4).

automatically prevent a decision from being classified as solely automated.[448] Clear instructions and training are needed for human decision-makers to avoid deferring blindly to AI outputs, especially given the risks of automation bias.

This conceptual and legal divergence is further reinforced by how each instrument defines its regulated actors. Under the GDPR, controllers bear the primary responsibilities for ensuring lawful data processing, including in ADM. Under the AI Act, regulatory obligations are assigned to providers and deployers.[449] However, providers under the AI Act are unlikely to qualify as controllers under the GDPR during the deployment phase, while deployers under the AI Act will typically be the controllers responsible for any resulting ADM.[450] In the development and testing phases, by contrast, providers are more likely to act as controllers under the GDPR, particularly when processing personal data for training, validation, or post-market monitoring as required under Articles 10 and 61 AI Act.[451]

5.1.2.2 Implications for Overlap and Applicability

Understanding the divergence between ADM and High-risk AI is crucial for evaluating the extent to which the AI Act addresses gaps in the existing oversight framework. The relationship between both regimes can be mapped across three primary scenarios.

First, there are AI systems that qualify both as high-risk and as ADM. In these cases, both legal frameworks apply. The AI Act imposes technical and organisational requirements on providers and developers. However, these requirements do not automatically guarantee compliance with the GDPR. A possible example is the SCHUFA case, in which the CJEU found that the use of a credit score generated by a traditional software system would fall under Article 22 GDPR.[452] Such a system could also fall under the high-risk obligations of that framework, including the requirements for human oversight set out in Article 14 AI Act.[453] In such cases, both regimes apply in parallel.

Second, there are cases where ADM systems do not fall within the scope of the AI Act's high-risk classification. These systems may still trigger Article 22 GDPR if they meet the conditions of being solely automated and producing significant effects. In such cases, the AI Act provides no additional safeguards. The example of recommender and content moderation systems, or behavioural

[448] CNPD, Deliberação n.º 2021/622, May 11, 2021; Rechtbank Den Haag, Case C/09/585239 / HA ZA 19-1221, ECLI:NL:RBDHA:2020:865 (11 February 2020).
https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1013>.
[449] As discussed in Section 4.2.3.
[450] Sebastião Barros Vale (n 4).
[451] EDPB, 'Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR' para 24.
[452] Case C-634/21 SCHUFA Holding [2023] ECLI:EU:C:2023:957.
[453] Such a system could arguably fall under Annex III, point 5(b), AI Act.

advertising tools, illustrates that some systems investigated by DPAs and courts under the GDPR have not been classified as a high-risk AI system, despite their potential for significant individual impact.

Third, there are systems that qualify as high-risk AI under the AI Act but do not perform ADM under the GDPR. This includes systems that generate recommendations, signals, or prompts used by human decision-makers, but where no fully automated decision is made. Although Article 14 AI Act requires that such systems be equipped with human oversight features, Article 22 GDPR may not apply at all. A further category of concern arises where neither the GDPR nor the AI Act applies effectively, especially when software tools do not rely on personal data but still produce outputs with significant real-world consequences.[454] An illustrative example is the use of computer-generated evidence in criminal proceedings, such as the Horizon accounting software used by the UK Post Office, which led to widespread miscarriages of justice.[455] Here, no personal data may be involved, and no automated decision in the sense of Article 22 GDPR is made, yet individuals are profoundly affected by the outputs.

Understanding these configurations is critical to determining whether the AI Act effectively reinforces human oversight protections or whether it leaves important gaps in the GDPR framework unaddressed. It also highlights the need for careful coordination between both instruments, particularly in terms of compliance responsibilities, actor definitions, and the conceptualisation of oversight.

5.2 Areas Where the AI Act Addresses GDPR Limitations

5.2.1 Proactive Design and Governance Duties

One of the most significant ways the AI Act attempts to overcome the GDPR's limitations is by shifting oversight obligations from reactive, individual safeguards to ex ante structural requirements. Under the GDPR, the architecture of ADM systems is largely left unregulated, with oversight framed as a remedial right rather than a design feature. The AI Act departs from this approach by embedding human oversight into the development process itself. It is explicitly required that high-risk AI systems be designed and developed to allow for effective human oversight throughout their lifecycle.[456] This obligation is reinforced by related provisions on transparency, deployer duties, and technical documentation, all of which help ensure that oversight is not merely formal but operationally feasible.

[454] Lisa McClory, 'Legal Requirements for Automated Decision-Making in the EU & UK' (*GLI*, 19 April 2024) <https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/profiling-and-prohibited-practices-a-summary-of-legal-requirements-for-automated-decision-making-in-the-eu-and-uk/> accessed 5 August 2025.
[455] Hamilton & others v Post Office Limited [2021] EWCA Crim 577
[456] Article 14(1) AI Act.

These provisions mark a move toward ensuring that oversight is not just theoretically possible, but technically feasible. Providers must now account for oversight at the system design stage, requiring them to anticipate the operational conditions under which humans can intervene meaningfully.[457] This represents a notable improvement over the GDPR, where oversight can be rendered impracticable by opaque or rigid system architecture.

However, this shift also introduces complexity. The Act does not specify how such design-stage considerations should be implemented in practice, nor does it clarify how much discretion providers have in defining what counts as 'effective' oversight. This opens the door to divergent interpretations and potential superficial compliance. The AI Act's reliance on providers' own risk management processes may thus fall short in ensuring that oversight capabilities are not only embedded but functionally operable.

Another important element of this shift is the introduction of the FRIA, required for deployers in the public sector and those providing services of general interest.[458] This instrument is intended to ensure that the use of high-risk AI systems is preceded by a contextual analysis of their potential impact on fundamental rights. It parallels the GDPR's DPIA, but adopts a broader lens by considering not only data protection risks but also social, economic, and non-discriminatory effects.[459] While its scope is more limited, applying only to certain deployers, the FRIA reflects the AI Act's attempt to institutionalise oversight at an earlier stage in the system lifecycle.[460] However, its effectiveness remains uncertain. The AI Act does not set out specific sanctions for non-compliance, instead leaving it to Member States to define penalties.[461] This creates a risk of uneven enforcement, particularly if public authorities are exempted from sanction regimes, as has occurred under the GDPR.[462] Without robust accountability, the FRIA risks becoming another formal requirement with limited preventive effect.

5.2.2 Clarified Oversight Roles and Implementation Requirements

In contrast to the GDPR's silence on who should exercise oversight and under what conditions, the AI Act introduces specific responsibilities for both providers and deployers.[463] This reflects a departure

[457] Article 14(3) AI Act.
[458] Article 27(1) AI Act.
[459] ibid; Levitina (n 56) 134–135.
[460] Alessandro Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI* (Springer Nature 2022) 48–51 <https://library.oapen.org/handle/20.500.12657/57009> accessed 7 August 2025.
[461] Article 99 AI Act.
[462] Mantelero (n 438) 10.
[463] As discussed in Section 4.2.3.

from the GDPR's vague reliance on human intervention, which, as Section 3.2.1.3 showed, often resulted in symbolic or unqualified involvement.

This clarity helps address the organisational ambiguity identified under the GDPR. Whereas the latter fails to integrate oversight into internal governance structures, the AI Act explicitly distributes obligations across the AI lifecycle. Deployers are responsible for ensuring that those tasked with oversight understand the system's functioning and can intervene appropriately. Providers must, in turn, offer adequate documentation and instructions for use, facilitating effective collaboration between both actors.

Nevertheless, implementation challenges persist. The requirement of sufficient competence remains underdefined, and no concrete standard is offered for evaluating the quality of oversight in practice.[464] Moreover, there is a risk that these provisions will not be accompanied by mechanisms to ensure meaningful accountability, such as audits or performance assessments. Without such supporting structures, the allocation of oversight roles may improve formal clarity but fall short of delivering substantive protection.

The AI Act's introduction of logging obligations also enhances traceability across the AI lifecycle, potentially strengthening the organisational conditions for effective oversight. By requiring providers and, in some cases, deployers to maintain detailed records of system operation, these provisions create a documented trail that could facilitate post-hoc review, error detection, and regulatory investigation.[465] However, the utility of such traceability for the data subject remains uncertain. Logging is primarily oriented toward internal governance and external supervisory authorities, rather than toward enabling individuals to understand or contest specific decisions affecting them. In the absence of complementary disclosure duties or accessible explanations, the mere existence of records may contribute little to the data subject's ability to exercise meaningful control.

5.2.3 From Reactive Remedies to Continuous Supervision

The AI Act also reorients the logic of human oversight by recognising that intervention is not merely a response to individual decisions but a continuous responsibility throughout the AI system's operation. This stands in contrast to the GDPR's more reactive safeguards, which primarily centre on the individual's right to contest decisions after harm has occurred.

[464] Levitina (n 56) 134; Crootof, Kaminski and W. Price II (n 41) 448.
[465] As discussed in Section 4.3.1.2.

As laid out before, it requires that human overseers be able to intervene in or interrupt the functioning of a high-risk system, implying an ongoing supervisory role rather than a one-time checkpoint.[466] Oversight, in this context, involves not only the ability to halt the system's operation, but also the responsibility to monitor its behaviour over time. This includes reviewing training data and algorithmic logic for bias, tracking performance for accuracy and unintended outcomes, and intervening in decision-making processes that may have significant impact.[467]

Additionally, the technical documentation requirements establish a framework for lifecycle oversight, including post-market monitoring and incident reporting.[468] These provisions promote a model of oversight grounded in proactive risk detection and iterative improvement.

This continuous orientation addresses a core limitation of the GDPR: the absence of any mandate to monitor whether oversight functions meaningfully after deployment. In the GDPR, the burden falls on individuals to detect failures and invoke their rights. In contrast, the AI Act suggests a structural model where oversight is embedded in organisational processes and accountability mechanisms.

Yet, here too, caution is warranted. The mere presence of lifecycle documentation and monitoring does not guarantee effective oversight unless these processes are integrated into institutional practice.

5.3 Remaining Gaps and the Continuing Role of the GDPR

5.3.1 Areas Where the AI Act Remains Silent

5.3.1.1 Inattention to Group-Level Harms and Societal Impact

While the AI Act introduces a dedicated oversight framework for high-risk systems, its protective scope remains narrowly confined to individual rights and safety. It lacks any clear legal mechanism for identifying or mitigating broader societal harm. This limited risk framing does not account for how AI may reinforce structural inequalities, deepen group-based discrimination, or destabilise public trust in democratic institutions. Such concerns cannot always be captured through the lens of personal data or individual redress.

A more comprehensive understanding of AI-related harm distinguishes three categories.[469] First, individual harm, which concerns adverse effects suffered by a specific person, such as discriminatory

[466] Article 14(4)(e) AI Act.
[467] Data Protection Authority of Belgium, 'Artificial Intelligence Systems and the GDPR: A Data Protection Perspective' (2024) 11–12 <https://www.autoriteprotectiondonnees.be/publications/artificial-intelligence-systems-and-the-gdpr---a-data-protection-perspective.pdf>.
[468] Article 11, Annex IV, and Articles 72-73 AI Act.
[469] Nathalie A. Smuha, 'Beyond the Individual: Governing AI's Societal Harm' (2021) 10 Internet Policy Review 4–5.

treatment or denial of services.[470] Second, collective harm, where identifiable groups, such as ethnic minorities or low-income communities, experience discriminatory patterns or structural exclusion resulting from algorithmic deployment. Third, societal harm, which cannot be reduced to either of the former and concerns setbacks to general societal interests, such as the erosion of transparency, democratic participation, or the rule of law. Unlike individual or group-based harm, societal harm often unfolds over time, through widespread or systemic use, and may not directly affect any one individual in a legally actionable way.[471]

While the AI Act introduces enforcement mechanisms and a public database for high-risk systems, these tools remain compliance-oriented and inaccessible to civil society. In contrast, the GDPR offers certain tools, most notably the right of access, that can in theory serve not only individual data protection but also broader public interest objectives.[472] When exercised collectively by civil society actors, this right has the potential to uncover discriminatory patterns or structural imbalances in ADM.[473] However, the uptake of this possibility has so far been limited, and the GDPR still lacks an explicit framework for addressing societal-level harms arising from ADM.

5.3.1.2 Challenges of Enforcement and Symbolic Compliance

Despite its structural ambitions, the AI Act does not fully escape the pitfalls of symbolic oversight.[474] The regulation's detailed role allocations, training requirements, and logging obligations will have limited effect if they are implemented as procedural formalities rather than as genuinely critical review processes. As with the GDPR, there is a risk that oversight becomes a mere procedural step in system deployment, serving as a 'rubber stamp' that legitimises flawed ADM outputs instead of meaningfully scrutinising them.[475] Without strong and systematic monitoring, the quality of human intervention may remain largely unexamined, allowing organisations to claim compliance by documenting processes that exist only in form.[476]

The GDPR faces a similar danger, but its reliance on individual rights of access, contestation, and remedy at least provides an avenue for those directly affected to trigger review and challenge unlawful processing. By contrast, the AI Act's enforcement logic is primarily mediated through governance and market surveillance authorities. While this may strengthen ex ante control at the system level, it also

[470] These interests may take various forms, including but not limited to physical, mental, economic, social, or legal dimensions; see ibid 5.
[471] ibid 9–12.
[472] René Louis Pierre Mahieu and Jef Ausloos, 'Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access' 16.
[473] See Mahieu and Ausloos (n 472).
[474] Identified in Section 3.2.2.1.
[475] Veale and Edwards (n 126) 400.
[476] Levitina (n 56) 137.

risks detaching oversight from those most affected by algorithmic decisions, who may never see the internal records or assessments that underpin compliance claims. Absent rigorous supervisory audits, performance assessments, and sanctions for inadequate oversight, the AI Act could replicate the same surface-level compliance practices it was intended to complement.

5.3.2 Where the GDPR Continues to Play a Protective Role

5.3.2.1 Individual Rights and Substantive Safeguards

Although the GDPR's reactive logic has limitations, it remains essential for ensuring transparency, contestability, and fairness in ADM. It grants individuals the right to obtain human intervention, express their views, and contest decisions, complemented by access rights under Article 15 GDPR.[477] These protections remain applicable even where the AI Act does not apply, including to systems that fall outside the high-risk classification or do not involve an 'AI system' as defined in Article 3 of the AI Act.[478]

Moreover, the GDPR embeds these rights within a broader legal framework that includes principles of fairness, data minimisation, and purpose limitation. These normative anchors are absent from the AI Act, which focuses primarily on technical risk mitigation. In this sense, the GDPR continues to provide substantive safeguards that the AI Act does not replicate, particularly in cases where personal data processing is central to the ADM system's functioning.

5.3.2.2 Regulatory Layering and Complementarity

In practice, many ADM systems will fall under both the GDPR and the AI Act. Their coexistence creates a layered regulatory environment that, when interpreted in a coordinated manner, can offer enhanced protection. For example, the AI Act may ensure structural readiness for oversight by imposing ex ante obligations on providers and deployers, while the GDPR complements this with individual rights to access, contest, and seek redress.

This layering is not merely theoretical. The AI Act itself recognises its relationship with the GDPR. Article 13 of the AI Act, for example, requires providers to supply information to deployers that is explicitly intended to support the conduct of a DPIA under Article 35 of the GDPR.[479] Similarly, the FRIA, may complement or even strengthen DPIA processes by broadening the assessment beyond personal data protection to include a wider range of fundamental rights.[480]

[477] Article 22(3) GDPR.
[478] Particularly given the concerns that certain systems may avoid being classified as high-risk by relying on the exceptions outlined in Section 4.1.2.2.
[479] Article 26(9) AI Act.
[480] Article 27(4) AI Act.

These forms of regulatory interdependence illustrate the potential for complementarity. However, this potential will only be realised if enforcement bodies coordinate their guidance and oversight. Divergent interpretations between data protection authorities and AI regulators, or inconsistent implementation at national level, risk creating legal uncertainty or fragmented protections. The success of this model therefore hinges not only on legal design, but on practical cooperation, institutional dialogue, and the development of shared interpretative standards.

If treated in isolation, the two regimes may produce gaps or duplication. But if understood as part of a mutually reinforcing framework, their interaction can support a more comprehensive system of safeguards, which combines structural preparedness with enforceable rights and continuous oversight across the AI lifecycle.

6 Conclusion

This thesis has examined the role of human oversight in EU law by analysing its current legal basis under the GDPR and its reformulation under the AI Act. Through this comparative and doctrinal inquiry, it has sought to answer the overarching research question: How does the AI Act reinforce, or diverge from the GDPR's framework for human oversight?

The analysis in Chapter 3 has shown that the GDPR, although often invoked in discussions on automated decision-making, provides only a limited legal framework for human oversight. Article 22 of the GDPR establishes a high threshold of application, relying on the ambiguous criterion of "solely automated" processing with "legal or similarly significant" effects. Even where Article 22 GDPR applies, its safeguards have proven difficult to operationalise in practice. Furthermore, the broader structural principles of the GDPR, such as fairness, transparency, and accountability, offer support for human oversight but lack specificity and enforceability in the context of algorithmic systems. The analysis also showed that the GDPR's approach has been constrained not only by its reactive logic and formalistic thresholds but also by an idealised view of human capacities. Practical, organisational, and behavioural challenges, including automation bias, complexity, and fragmented accountability, undermine the effectiveness of human involvement. These findings underscore the need for a more structured, proactive, and risk-sensitive oversight framework that reflects the realities of modern AI systems.

Against this background, Chapter 4 examined how the AI Act introduces new provisions that aim to address these shortcomings. Unlike the GDPR, the AI Act applies a preventive logic rooted in risk-based regulation. Oversight obligations are reserved for high-risk AI systems, and Article 14 AI Act requires both design-based and organisational safeguards to ensure that oversight is effective, timely, and meaningful. The AI Act shifts the focus from ex post human intervention to ex ante responsibility. It embeds oversight capabilities at the development stage and clarifies the respective roles of providers and deployers. Complementary provisions further support this framework by requiring documentation, traceability, and instructions for use that are intelligible to human overseers.

However, as Chapter 5 has shown, the extent to which the AI Act resolves the GDPR's oversight deficits remains open. While the AI Act introduces clearer and more prescriptive obligations, several challenges persist. Key concepts such as 'effective' oversight remain undefined, and implementation is left to be determined largely by providers and deployers. Moreover, oversight is still shaped by the high-risk classification, which leaves many impactful systems outside the scope of mandatory obligations. The risk of formalistic compliance therefore remains. This risk is especially present where organisational and contextual factors are not adequately addressed. The AI Act marks a regulatory

shift, but its practical impact will depend on how its provisions are interpreted, enforced, and complemented by future guidance or case law.

These findings resonate with wider academic debates on algorithmic governance and human oversight. Scholars have long questioned whether individual rights-based safeguards can adequately address the systemic risks of automated decision-making and have criticised the GDPR for relying on reactive, post-hoc remedies. Others have explored how oversight should be embedded in organisational design and supported by technical and institutional guarantees. By systematically comparing the GDPR and the AI Act, this thesis contributes to these debates by showing how the EU is beginning to move from an individual rights model towards a more proactive, risk-based architecture of oversight, while also exposing where that transition remains incomplete.

In conclusion, the AI Act represents a significant development in the legal architecture of human oversight. It builds on the GDPR's foundations while introducing more targeted and proactive mechanisms. Navigating these frameworks effectively requires constant attention to their points of divergence and overlap, particularly between high-risk AI under the AI Act and ADM under the GDPR, as this determines when obligations arise and how they interact in practice. Yet the transformation is neither complete nor guaranteed. The real test lies ahead. As I cannot predict how these frameworks will evolve in practice, only time will tell whether the AI Act will truly mark a turning point. Much will depend on how oversight is implemented on the ground, how ambiguous terms are clarified through jurisprudence and standards, and whether institutional actors will ensure that oversight is not merely present but meaningful. As such, the future of human oversight in the age of AI remains a dynamic and contested space. It is one that demands continued legal, technical, and ethical scrutiny.

7 Bibliography

7.1 Legislation

Consolidated version of the Treaty on European Union [2012] OJ C326/13

Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C326/47

Charter of Fundamental Rights of the European Union [2012] OJ C326/391

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 ('DPD')

Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés

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 [2016] OJ L119/1 ('GDPR')

Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision [2016] OJ L119/89 ('LED')

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [2022] OJ L227/1 ('DSA')

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L2024/1689 ('AI Act')

Wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens, BS 3 februari 1999

7.2 Case-law

7.2.1 Court of Justice

Case C-203/22 Dun & Bradstreet Austria GmbH ECLI:EU:C:2024:745, Opinion of AG De La Tour

Case C-203/22 Dun & Bradstreet Austria GmbH ECLI:EU:C:2025:117

Case C-634/21 Case C-634/21 SCHUFA Holding [2023] ECLI:EU:C:2023:957

7.2.2 National Courts

Bundesgerichtshof, VI ZR 156/13 (28 January 2014) (Germany).

Rechtbank Amsterdam, C/13/687315 / HA RK 20-207, ECLI:NL:RBAMS:2021:1020 (11 March 2021) (Netherlands)

Rechtbank Amsterdam, C/13/689705 / HA RK 20-258, ECLI:NL:RBAMS:2021:1019 (11 March 2021) (Netherlands)

Rechtbank Amsterdam, C/13/692003 / HA RK 20-302, ECLI:NL:RBAMS:2021:1018 (11 March 2021) (Netherlands)

Rechtbank Amsterdam, 742407 / HA RK 23-366, ECLI:NL:RBAMS:2024:4019 (4 July 2024) (Netherlands)

Rechtbank Den Haag, C/09/550982 / HA ZA 18-388, ECLI:NL:RBDHA:2020:1878 (5 February 2020) (Netherlands)

Rechtbank Den Haag, C/09/585239 / HA ZA 19-1221, ECLI:NL:RBDHA:2020:865 (11 February 2020) (Netherlands)

Tribunal Administratif de Marseille, La Quadrature du Net, No 1901249 (27 November 2020) (France)

Ústavný súd Slovenskej republiky (Constitutional Court of Slovakia), Case 492/2021 Z. z. (10 November 2021) (Slovakia)

Hamilton and others v Post Office Limited [2021] EWCA Crim 577 (United Kingdom)

State v Loomis, Supreme Court of Wisconsin (United States)

7.3 Doctrine

Alexander V, Blinder C and Zak PJ, 'Why Trust an Algorithm? Performance, Cognition, and Neurophysiology' (2018) 89 Computers in Human Behavior 279 <https://www.sciencedirect.com/science/article/pii/S0747563218303480> accessed 5 August 2025

Almada, Marco, 'Automated Decision-Making as a Data Protection Issue' [2021] Social Science Research Network <https://www.researchgate.net/publication/350759609_Automated_decision-making_as_a_data_protection_issue> accessed 8 August 2025

Anderson B, 'Tesla Says It's Driverless But Someone's Always Watching' (Carscoops, 23 June 2025) <https://www.carscoops.com/2025/06/tesla-robotaxis-arrive-in-austin-but-its-using-safety-drivers/> accessed 22 July 2025

Bainbridge L, 'Ironies of Automation' (1983) 19 Automatica 775 <https://www.sciencedirect.com/science/article/pii/0005109883900468> accessed 3 August 2025

Bauer K and others, 'Expl(AI)n It to Me – Explainable AI and Information Systems Research' (2021) 63 Business & Information Systems Engineering 79 <https://doi.org/10.1007/s12599-021-00683-2> accessed 8 August 2025

Bertolini A, 'Artificial Intelligence and Civil Liability' (Policy Department for Justice, Civil Liberties and Institutional Affairs 2025) <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/776426/IUST_STU(2025)776426_EN.pdf>

Bhabendu Kumar Mohanta, Soumyashree S Panda, and Debasish Jena, 'An Overview of Smart Contract and Use Cases in Blockchain Technology', ResearchGate (2025) <https://www.researchgate.net/publication/328581609_An_Overview_of_Smart_Contract_and_Use_Cases_in_Blockchain_Technology> accessed 22 July 2025

Binns R and Veale M, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR' (2021) 11 International Data Privacy Law 319 <https://doi.org/10.1093/idpl/ipab020> accessed 26 July 2025

Brkan M, 'Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond' (Social Science Research Network, 1 August 2017) <https://papers.ssrn.com/abstract=3124901> accessed 29 June 2025

Burrell J, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 2053951715622512 <https://doi.org/10.1177/2053951715622512> accessed 28 June 2025

Busuioc M, Curtin D and Almada M, 'Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act' (2023) 2 European Law Open 79 <https://www.cambridge.org/core/journals/european-law-open/article/reclaiming-transparency-contesting-the-logics-of-secrecy-within-the-ai-act/01B90DB4D042204EED7C4EEF6EEBE7EA> accessed 29 July 2025

Cobbe J, Lee MSA and Singh J, 'Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021) <https://dl.acm.org/doi/10.1145/3442188.3445921> accessed 8 August 2025

Crootof R, Kaminski M and W. Price II, 'Humans in the Loop' (2023) 76 Vanderbilt Law Review 429 <https://scholarship.law.vanderbilt.edu/vlr/vol76/iss2/2>

Ebers M, 'Truly Risk-Based Regulation of Artificial Intelligence - How to Implement the EU's AI Act' (Social Science Research Network, 19 June 2024) <https://papers.ssrn.com/abstract=4870387> accessed 24 July 2025

——, 'AI Robotics in Healthcare Between the EU Medical Device Regulation and the Artificial Intelligence Act' (2024) 11 Oslo Law Review 1 <https://www.scup.com/doi/10.18261/olr.11.1.2> accessed 25 July 2025

Edwards L, 'Regulating AI in Europe: Four Problems and Four Solutions (Ada Lovelace Institute Expert Opinion)' (Social Science Research Network, 1 March 2022) <https://papers.ssrn.com/abstract=5026691> accessed 21 July 2025

Elish MC, 'Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (Pre-Print)' (Social Science Research Network, 1 March 2019) <https://papers.ssrn.com/abstract=2757236> accessed 3 August 2025

Enarsson T, Enqvist L and Naarttijärvi M, 'Approaching the Human in the Loop – Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts' (2022) 31 Information & Communications Technology Law 123 <https://doi.org/10.1080/13600834.2021.1958860> accessed 28 April 2024

Enqvist L, '"Human Oversight" in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15 Law, Innovation and Technology 508 <https://doi.org/10.1080/17579961.2023.2245683> accessed 22 December 2024

Ferraris V and others, 'Defining Profiling' (Social Science Research Network, 11 December 2013) <https://papers.ssrn.com/abstract=2366564> accessed 22 July 2025

Fink M, 'Human Oversight under Article 14 of the EU AI Act' (Social Science Research Network, 14 February 2025) <https://papers.ssrn.com/abstract=5147196> accessed 25 June 2025

Fischhoff B, Watson SR and Hope C, 'Defining Risk' (1984) 17 Policy Sciences 123 <https://doi.org/10.1007/BF00146924> accessed 25 July 2025

Frank Knight, *Risk, Uncertainty, and Profit* (1921) <https://fraser.stlouisfed.org/files/docs/publications/books/risk/riskuncertaintyprofit.pdf> accessed 11 August 2025

Gil González E and De Hert P, 'Understanding the Legal Provisions That Allow Processing and Profiling of Personal Data—an Analysis of GDPR Provisions and Principles' (2019) 19 ERA Forum 597 <http://link.springer.com/10.1007/s12027-018-0546-z> accessed 23 June 2025

Green B, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 Computer Law & Security Review 105681 <https://www.sciencedirect.com/science/article/pii/S0267364922000292> accessed 28 April 2024

Green B and Chen Y, 'Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts' (2021) 5 Proc. ACM Hum.-Comput. Interact. 418:1 <https://dl.acm.org/doi/10.1145/3479562> accessed 3 August 2025

Gültekin-Várkonyi DG, 'AI Literacy for Legal AI Systems: A Practical Approach' (Social Science Research Network, 20 May 2025) <https://papers.ssrn.com/abstract=5309725> accessed 21 July 2025

Harasimiuk D and Braun T, *Regulating Artificial Intelligence: Binary Ethics and the Law* (Routledge 2021)

Heather Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>

Hildebrandt M, 'The Disconnect Between "Upstream" Automation and Legal Protection Against Automated Decision Making' (*Technology Law*, 7 April 2022) <https://cyber.jotwell.com/the-disconnect-between-upstream-automation-and-legal-protection-against-automated-decision-making/> accessed 26 July 2025

Holzinger A, Zatloukal K and Müller H, 'Is Human Oversight to AI Systems Still Possible?' (2025) 85 New Biotechnology 59 <https://www.sciencedirect.com/science/article/pii/S1871678424005636> accessed 25 June 2025

Kaminski ME and Malgieri G, 'The Right to Explanation in the AI Act' (Social Science Research Network, 8 March 2025) <https://papers.ssrn.com/abstract=5194301> accessed 7 July 2025

Koen Smit and Martijn Zoet, 'A Governance Framework for (Semi) Automated Decision-Making', *Proceedings of the Tenth International Conference on Information, Process, and Knowledge Management* (2018) <https://www.researchgate.net/publication/325975696_A_Governance_Framework_for_Semi_Automated_Decision-Making> accessed 5 August 2025

Kuner C and others, 'The EU General Data Protection Regulation: A Commentary/Update of Selected Articles' [2021] SSRN Electronic Journal <https://www.ssrn.com/abstract=3839645> accessed 27 April 2024

Laux J, 'Institutionalised Distrust and Human Oversight of Artificial Intelligence: Towards a Democratic Design of AI Governance under the European Union AI Act' (2024) 39 AI & SOCIETY 2853 <https://doi.org/10.1007/s00146-023-01777-z> accessed 25 June 2025

Lazcoz G and de Hert P, 'Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities' (2023) 50 Computer Law & Security Review 105833 <https://www.sciencedirect.com/science/article/pii/S0267364923000432> accessed 15 March 2025

Levitina A, 'Humans in Automated Decision-Making under the GDPR and AI Act' [2024] Revista CIDOB d'Afers Internacionals 121 <https://cidob.org/en/publications/humans-automated-decision-making-under-gdpr-and-ai-act> accessed 26 July 2025

Mahieu RLP and Ausloos J, 'Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access'

Malgieri G, '"Just" Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation' (2021) 1 Law and Business 16 <https://sciendo.com/article/10.2478/law-2021-0003> accessed 8 August 2025

Malgieri G and Comandé G, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (Social Science Research Network, 13 November 2017) <https://papers.ssrn.com/abstract=3088976> accessed 24 June 2025

Malgieri G and Pasquale F, 'Licensing High-Risk Artificial Intelligence: Toward Ex Ante Justification for a Disruptive Technology' (2024) 52 Computer Law & Security Review 105899 <https://www.sciencedirect.com/science/article/pii/S0267364923001097> accessed 4 July 2025

Mantelero A, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI* (Springer Nature 2022) <https://library.oapen.org/handle/20.500.12657/57009> accessed 7 August 2025

——, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 Computer Law & Security Review 106020 <http://arxiv.org/abs/2411.15149> accessed 1 August 2025

McCorduck P and Cfe C, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (2nd edn, A K Peters/CRC Press 2004)

Mendoza I and Bygrave LA, 'The Right Not to Be Subject to Automated Decisions Based on Profiling' (Social Science Research Network, 8 May 2017) <https://papers.ssrn.com/abstract=2964855> accessed 26 June 2025

Nathalie A. Smuha, 'Beyond the Individual: Governing AI's Societal Harm' (2021) 10 Internet Policy Review

Novelli C and others, 'Taking AI Risks Seriously: A New Assessment Model for the AI Act' (2024) 39 AI & SOCIETY 2493 <https://doi.org/10.1007/s00146-023-01723-z> accessed 24 July 2025

Okpala I, Golgoon A and Kannan AR, 'Agentic AI Systems Applied to Tasks in Financial Services: Modeling and Model Risk Management Crews' (arXiv, 29 April 2025) <http://arxiv.org/abs/2502.05439> accessed 22 July 2025

Parasuraman R and Manzey DH, 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52 Human Factors 381 <https://doi.org/10.1177/0018720810376055> accessed 24 June 2025

Pasquale F, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015) <https://www.jstor.org/stable/j.ctt13x0hch> accessed 7 August 2025

Rai A, Constantinides P and Sarker S, 'Next Generation Digital Platforms : Toward Human-AI Hybrids' (2019) 43 MIS Quarterly iii <https://misq.org/misq/downloads/> accessed 21 July 2025

Riikka Koulu, 'Human Control over Automation : EU Policy and AI Ethics' (2020) 12 European journal of legal studies <https://hdl.handle.net/1814/66992>

Ruschemeier H, 'AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23 ERA Forum 361 <https://doi.org/10.1007/s12027-022-00725-6> accessed 26 July 2025

Schmidt P, Biessmann F and Teubner T, 'Transparency and Trust in Artificial Intelligence Systems' (2020) 29 Journal of Decision Systems 260 <https://doi.org/10.1080/12460125.2020.1819094> accessed 8 August 2025

Schwemer SF, Tomada L and Pasini T, 'Legal AI Systems in the EU's Proposed Artificial Intelligence Act' (Social Science Research Network, 21 June 2021) <https://papers.ssrn.com/abstract=3871099> accessed 1 August 2025

Selbst AD and Barocas S, 'The Intuitive Appeal of Explainable Machines' (Social Science Research Network, 2 March 2018) <https://papers.ssrn.com/abstract=3126971> accessed 24 June 2025

Skitka LJ, Mosier KL and Burdick M, 'Does Automation Bias Decision-Making?' (1999) 51 International Journal of Human-Computer Studies 991 <https://linkinghub.elsevier.com/retrieve/pii/S1071581999902525> accessed 11 December 2024

Stephan Dreyer and Wolfgang Schulz, 'The General Data Protection Regulationand Automated Decision-Making:Will It Deliver?' [2019] Bertelsmann Foundation

Sterz S and others, 'On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives', *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2024) <https://dl.acm.org/doi/10.1145/3630106.3659051> accessed 22 December 2024

Stuart Russel and Peter Norvig, *Artificial Intelligence A Modern Approach* (Fourth Edition Global Edition, Pearson 2021)

Sweeney L, ‘Discrimination in Online Ad Delivery’ (2013) 56 Commun. ACM 44 <https://dl.acm.org/doi/10.1145/2447976.2447990> accessed 2 August 2025

Teresa Rodríguez de las Heras Ballell, ‘Guiding Principles for Automated Decision-Making in the EU’ (2022) ELI Innovation Paper

Tiago Sérgio Cabral, ‘AI and the Right to Explanation: Three Legal Bases under the GDPR, *Data Protection and Privacy* (2021) <https://www.researchgate.net/publication/349496743_AI_and_the_Right_to_Explanation_Three_Legal_Bases_under_the_GDPR> accessed 30 June 2025

Tosoni L, ‘The Right To Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation’ (Social Science Research Network, 14 May 2021) <https://papers.ssrn.com/abstract=3845913> accessed 2 July 2025

Turing AM, ‘Computing Machinery and Intelligence’ (1950) LIX Mind 433 <https://doi.org/10.1093/mind/LIX.236.433> accessed 21 July 2025

Veale M and Borgesius FZ, ‘Demystifying the Draft EU Artificial Intelligence Act’ (2021) 22 Computer Law Review International 97 <http://arxiv.org/abs/2107.03721> accessed 24 December 2024

Veale M and Edwards L, ‘Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling’ (2018) 34 Computer Law & Security Review 398 <https://www.sciencedirect.com/science/article/pii/S026736491730376X> accessed 1 July 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://doi.org/10.1093/idpl/ipx005> accessed 28 April 2024

‘Xenophobic Machines: Discrimination through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal’ (*Amnesty International*, 25 October 2021) <https://www.amnesty.org/en/documents/eur35/4686/2021/en/> accessed 24 June 2025

Zuiderveen Borgesius F and Poort J, ‘Online Price Discrimination and EU Data Privacy Law’ (2017) 40 Journal of Consumer Policy 347 <https://doi.org/10.1007/s10603-017-9354-z> accessed 2 August 2025

7.4 Other

Agencia Española de Protección de Datos (AEPD), Procedimiento Nº: E/03624/2021(2021) <https://www.aepd.es/es/documento/e-03624-2021.pdf>

'AI Act | Shaping Europe's Digital Future' (24 July 2025) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 24 July 2025

'AI Watch: Global Regulatory Tracker - China | White & Case LLP' (29 May 2025) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china> accessed 21 September 2025

Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data 1992

Anderson B, 'Tesla Says It's Driverless But Someone's Always Watching' (*Carscoops*, 23 June 2025) <https://www.carscoops.com/2025/06/tesla-robotaxis-arrive-in-austin-but-its-using-safety-drivers/> accessed 22 July 2025

Article 29 Data Protection Working Party, 'Opinion 02/2013 on Apps on Smart Devices' <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf>

Autoriteit Persoonsgegevens, 'Boete Belastingdienst kinderopvangtoeslag' <https://www.autoriteitpersoonsgegevens.nl/documenten/boete-belastingdienst-kinderopvangtoeslag> accessed 4 July 2025

CNPD, Deliberação n.º 2021/622, May 11, 2021; Rechtbank Den Haag, Case C/09/585239 / HA ZA 19-1221, ECLI:NL:RBDHA:2020:865 (11 February 2020)

Data Protection Authority of Belgium, 'Artificial Intelligence Systems and the GDPR: A Data Protection Perspective' (2024) <https://www.autoriteprotectiondonnees.be/publications/artificial-intelligence-systems-and-the-gdpr---a-data-protection-perspective.pdf>

'Dutch Scandal Serves as a Warning for Europe over Risks of Using Algorithms' *POLITICO* (29 March 2022) <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/> accessed 24 June 2025

EDPB, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679' (2017) <https://ec.europa.eu/newsroom/article29/items/611236> accessed 6 July 2025

——, 'Guidelines 01/2022 on Data Subject Rights - Right of Access' (2023) <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-012022-data-subject-rights-right-access_en> accessed 6 July 2025

——, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en> accessed 2 August 2025

——, 'Guidelines 05/2020 on Consent under Regulation 2016/679' <https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en> accessed 28 June 2025

——, ‘Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR’

‘EDPB-EDPS Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)’

EDPB/WP29, ‘Article 29 Working Party Guidelines on Transparency under Regulation 2016/679’ (2018) <https://www.edpb.europa.eu/system/files/2023-09/wp260rev01_en.pdf>

——, ‘Article 29 Data Protection Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, WP251rev.01’ (2018) <https://ec.europa.eu/newsroom/article29/items/612053/en> accessed 13 March 2025

European Commission, ‘Impact Assessment of the Regulation on Artificial Intelligence’ (2021)

European Commission, Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) COM(2021) 206 final

European Commission, ‘White Paper on Artificial Intelligence – A European Approach to Excellence and Trust’ COM(2020) 65 final

Garante per la protezione dei dati personali, Ordinanza ingiunzione nei confronti di Deliveroo Italy S.r.l. (22 July 2021) [9685994] <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9685994>

Gegevensbeschermingsautoriteit (GBA), Decision on the Merits No. 109/2024 (29 August 2024) <https://www.gegevensbeschermingsautoriteit.be/publications/beslissing-ten-gronde-nr.-109-2024.pdf>

Gianmarco Gori, Figure ‘Software, Data and AI Value Chain’ in Legal Professionals in the Digital Age.(Regulatory Framework and Capita Selecta) (Vrije Universiteit Brussel, 2024)

‘Guidance on the AI Auditing Framework’ (UK Information Commissioner’s Office 2020)

High-Level Expert Group on AI, ‘Ethics Guidelines for Trustworthy AI | Shaping Europe’s Digital Future’ (2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 22 December 2024

Hunt T, ‘How Artificial Intelligence Will Change Administrative Law: The Government of Canada’s Directive on automated Decision-Making | DLA Piper’ <https://www.dlapiper.com/en-us/insights/publications/2023/05/how-artificial-intelligence-will-change-administrative-law-in-canada> accessed 3 August 2025

IMY, Decision in case DI-2019-4062 concerning Klarna Bank AB (28 March 2022) https://www.imy.se/globalassets/dokument/beslut/klarna/beslut-tillsyn-klarna.pdf accessed 5 July 2025, 18

Kostiantyn Ponomarov, ‘Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview’ (25 August 2025) <https://legalnodes.com/article/global-ai-regulations-tracker> accessed 21 September 2025

Lisa McClory, ‘Legal Requirements for Automated Decision-Making in the EU & UK’ (*GLI*, 19 April 2024) <https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/profiling-and-prohibited-practices-a-summary-of-legal-requirements-for-automated-decision-making-in-the-eu-and-uk/> accessed 5 August 2025

‘Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration’ (European Law Institute)

‘MPNE Patient Consensus on Data and AI 2.0’ (*Issuu*, 18 October 2024) <https://issuu.com/mpne/docs/mpne_consensus_2.0_long> accessed 29 July 2025

Sebastião Barros Vale, ‘GDPR and the AI Act Interplay: Lessons from FPF’s ADM Case-Law Report - Future of Privacy Forum’ (*Future of Privacy Forum*, 2022) <https://fpf.org/blog/gdpr-and-the-ai-act-interplay-lessons-from-fpfs-adm-case-law-report/> accessed 5 August 2025

Sebastião Barros Vale and Gabriela Zanfir-Fortuna, ‘Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities’ (Future of Privacy Forum 2022) <https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf> accessed 3 July 2025

Tambiama Madiega, Artificial Intelligence Act (EU Legislation in Progress Briefing, European Parliamentary Research Service, PE 698.792, September 2024)

‘Xenophobic Machines: Discrimination through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal’ (*Amnesty International*, 25 October 2021) <https://www.amnesty.org/en/documents/eur35/4686/2021/en/> accessed 24 June 2025