JuLIA (101046631)
"Justice, fundamentaL rIghts and Artificial Intelligence"
DG Just – JUST-2021-JTRA

JuLIA Handbook

AI and Public Administration: The (legal) limits of algorithmic governance

THIS PUBLICATION IS FUNDED BY THE EUROPEAN UNION'S JUSTICE PROGRAMME (2021-2027)

Scientific Coordinator of the JuLIA Project:
Mireia Artigot Golobardes (University of Pompeu Fabra)

Project Manager:
Bruno Castro Coma (University of Pompeu Fabra)

Coordinator of the team of legal experts on AI and Public Administration:
Aurelia Colombi Ciacchi (University of Groningen)

Editors of this Handbook:
Aurelia Colombi Ciacchi, María Lorena Flórez Rojas, Lottie Lane, Tobias Nowak

Coauthors of this Handbook (in alphabetical order):

Aminu Ibrahim, Yusrah
Ancuceanu, Beatrice
Ardenghi, Linda
Berardi, Greta
Bularca, Larisa-Diana
Carbone, Roberto
Colombi Ciacchi, Aurelia
Davidsson, Olivia
De Mari Casareto dal Verme, Tommaso
Emanuilov, Ivo
Flórez Rojas, María Lorena
Haitsma, Lucas
Lane, Lottie
Marchesini, Eleonora
Molinari, Sara
Nowak, Tobias
Onaie, Alma
Ranise, Silvio
Stoyanova, Ginka
Tsvetkova, Alexandra
Varošanec, Ida

May 2025

Acknowledgments
We thank all JuLIA partners for sharing materials and inputs underlying the analysis developed in this Handbook and Sara Petraskova for her editing work.

Table of Contents

1. Introduction to the Handbook 5
    1.1 Problem definition 5
    1.2 Scope of the Handbook 6
        1.2.1 Defining artificial intelligence 7
        1.2.2 Defining 'public administration' 11
        1.2.3 Illustrating uses of AI in public administration 11
    1.3 Structure of this Handbook 18
    1.4 Methodology 20
2. Algorithmic Design 21
    2.1 Risks for fundamental rights in the design stage of AI systems used in public administration 22
    2.2 European legal frameworks and case law 23
        2.2.1 The European Union 24
        2.2.2 The Council of Europe 27
    2.3 National case law and legal frameworks: Examples from Europe and beyond 27
        2.3.1 Legislation 27
        2.3.2 National case law 28
    2.4 National and international AI strategies 33
        2.4.1 National AI strategies 33
        2.4.2 International level 36
    2.5 Ethical instruments and Soft-law/policy framework 37
    2.6 AI Liability In The Public Sector 40
        2.6.1 Civil AI Liability 41
        2.6.2 AI Liability Directive and Product Liability Directive 43
        2.6.3 Liability Gaps. AI as Actor vs AI as Tool 45
        2.6.4 Closing Gaps. Article 47 CFR 46
        2.6.5 Conclusion 47
    2.7. Concluding remarks: Algorithmic design stage 47
        2.7.1 Guiding questions 48
3. Organisational implementation/embedding 49
    3.1 Risks for fundamental rights in the implementation of AI systems used in public administration 49
    3.2 Implementation stage of AI life cycle 49

3.3 Fundamental rights issues in the implementation stage- Study cases of real applications 52
3.4 Legal, regulatory and ethical frameworks applicable to implementation of AI systems used in public administration 55
3.5 European legal framework and case law 56
3.5.1 Legal framework 56
3.5.2 Case law 63
3.6 Liability and effective remedies 74
3.6.1 Use of AI and Adverse Impact on Fundamental Rights 74
3.6.2 The need for remedies 75
3.6.3 Legal Liability and the Claim for Damages 76
3.6.4 Injunctions and Invalid Administrative Acts 78
3.7 Concluding remarks: Organisational embedding stage 81
3.7.1. Guiding questions 82
4. Public decision-making using AI systems 83
4.1 Risks for fundamental rights in the decision-making using AI systems 83
4.2 Decision-making stage of AI life cycle 84
4.3 Fundamental rights issues in the decision-making stage- Study cases of real applications 87
4.3.1. EU case law 87
4.3.2. National courts 90
4.4 Legal, regulatory and ethical frameworks applicable to decision-making use of AI systems in public administration 95
4.4.1 Legal framework 95
4.4.2 Ethical concerns relating to the use of AI systems in public decision-making 106
4.5 Concluding remarks: Decision-making stage 109
4.5.1 Guiding questions 110
5. Conclusions 112

1. Introduction to the Handbook

Lottie Lane, Lucas Haitsma, Tobias Nowak, Tommaso De Mari Casareto dal Verme, Eleonora Marchesini, Roberto Carbone, Silvio Ranise, Olivia Davidsson, Sara Molinari

The JuLIA Handbook on AI and Public Administration: The (legal) limits of algorithmic governance builds upon the collaborative venture developed in previous projects of judicial training and, more recently, in the FRICoRe project.[1] In continuity with previous projects, this collaboration combines rigorous methodologies with judicial practices and provides judicial trainers, lawyers, judges, and other legal experts alike with the sort of rich comparative material that should always characterise transnational training sessions. Training includes not only the transfer of knowledge, but also the creation of a learning community composed of different professional skills. Like in previous experiences, this Handbook is due to evolve both in content and in method over time, with additional suggestions arising from its use in training events.

1.1 Problem definition

Artificial intelligence (AI) applications are increasingly used in various aspects of public administration. This includes internal uses within governmental authorities, such as internal management, as well as external uses that affect individuals external to public authorities, such as in the context of public service provision and the detection of fraud in welfare systems. Whether for reasons of efficiency, speed, perceived accuracy or cost-effectiveness, politicians and public administrators expect a plethora of benefits from using AI. However, the use of AI in public administration often comes with serious risks to fundamental rights. By now, some uses of AI in public administration have become infamous. For example, the Dutch childcare benefits scandal in the Netherlands led to the resignation of the Dutch government after thousands of families were falsely accused of fraud based on the results of an algorithm with devastating and discriminatory effects.[2] In Wales (UK), the use of automated facial recognition technology by police without sufficient limitations was found to be a violation of the right to privacy.[3] The Court argued that the police were not able to assess the potential discriminatory nature of the technology.

Risks to several rights protected in the Charter of Fundamental Rights of the European Union (CFREU)[4] are highlighted in this Handbook. The right to non-discrimination (Article 21), respect for private and family life (Article 7) and data protection (Article 8) are just a few

[1] Funded by the Justice Programme of the European Union, FRICoRe was a three-year judicial training project providing judges and legal practitioners with guidelines on the choice of procedures and remedies in enforcing fundamental rights at national level, with particular reference to the areas of consumer protection, migration and asylum, data protection, health law and non-discrimination. See: www.fricore.eu.

[2] Tweede Kamer der Staten-Generaal: Verslag - Parlementaire ondervragingscommissie Kinderopvangtoeslag. Ongekend onrecht. (Kamerstuk 35 510, nr. 2). 2020. (Report of the Inquiry Commission on child benefits payments of the Dutch Parliament).

[3] R Bridges v CC South Wales [2020] EWCA Civ 1058.

[4] Charter of Fundamental Rights of the European Union (2000) C364/01.

examples of rights that have already been interfered with due to public sector reliance on AI. However, the proliferation of the use of AI in public administration, as well as the understanding that reliance on AI should be considered from a fundamental rights perspective, are relatively recent. As a result, although many situations of potential violations of fundamental rights have come to light in recent years, the number of legal cases based on fundamental rights arguments is very limited. In addition, the legal and regulatory frameworks applicable to AI and fundamental rights are still being developed. Both established and emerging EU legislation on data protection, consumer protection, liability and the regulation of AI more generally, are relevant here. Additionally, a huge number of international, European and national initiatives try to guide developers and users of AI systems to develop and use AI in a rights-friendly manner. This includes binding and non-binding legal instruments, policy documents, and self-regulatory measures.

All of this raises difficult questions for judges dealing with cases concerning the design, implementation and use of AI systems for public administration:

- What is the (potential) impact of such systems on fundamental rights?
- What is the applicable (EU) law and policy in such situations and how has it been applied in previous cases?
- What is the role of the CJEU in cases raising potential violations of fundamental rights?
- Where does liability lie when fundamental rights are negatively affected, especially given the different actors involved in the development, implementation and use of systems?
- How should judges approach cases involving AI and fundamental rights?

We address these questions in this Handbook. Through the use of real-life cases and situations across Europe, we:

- explain how such AI systems work and are used;
- introduce key risks that they can pose to fundamental rights;
- lay down applicable (emerging) law and policy frameworks; and
- provide guidance as to how judges can approach cases concerning AI in the context of public administration.

1.2 Scope of the Handbook

The use of AI in public administration covers a broad range of issues with a wide range of potential human rights impacts. In order to provide meaningful guidance to judges in this Handbook, we have restricted the scope of the topic concerning the definition of both artificial intelligence and public administration. The following paragraphs explain the scope of the Handbook and provide examples of uses of AI in public administration falling within this scope. This includes some discussion of the more technical aspects that are key to understanding the impact of AI systems.

1.2.1 Defining artificial intelligence

The term 'artificial intelligence' does not have a universally accepted definition but can cover a wide range of technology systems.[5] For the purposes of continuity, the JuLIA project replicate the definition of AI that is found in the draft EU regulation on AI – the AI Act (article 3 (1)):

> AI system' means a machine-based system that is designed to operate with varying levels of autonomy and 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.[6]

Although identifying a clear definition is a major challenge even within the scientific community, it is possible to understand the modern approach to AI research in opposition to what is to be intended as "traditional AI". Russel and Norvig, the main modern AI researchers, start from a central question: 'why would we want an agent to learn?'[7] They give three answers:

1. The designers cannot anticipate all possible situations that the agent might find itself in;
2. The designers cannot anticipate all changes over time;
3. Sometimes human programmers have no idea how to program a solution themselves.

Learning techniques represent the main features of "modern AI", which is capable of acting autonomously, learning from its own experience, evolving and improving its performance over time. This is possible today thanks to the implementation of machine learning (ML) and deep learning (DL) techniques combined with the huge development of computers' power and the global diffusion of the Internet.

Machine learning

Machine learning (ML) focuses on the development of algorithms and models that enable computers to learn from data and make predictions or decisions without explicit programming. The core idea is to allow machines to improve their performance on a specific task over time by learning from experience. It employs statistical theories in constructing

[5] See also S Ranise, Technical & risk analysis in the JuLIA Handbook on Artificial Intelligence, Judicial Decision-Making and Fundamental Rights and Chapter 1 section 1.2 of the JuLIA Handbook on Artificial intelligence in Healthcare and the Protection of Health as a Fundamental Right.

[6] 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 (Artificial Intelligence Act) [2024] OJ L 2024/1689 (AI Act).

[7] Russell, S.J. and P. Norvig. (2022) Artificial intelligence : a modern approach. Fourth edition. Global edition. Harlow, United Kingdom: Pearson.

mathematical models since the main task of the learning algorithm is to create deductions and conclusions from examples. To this end, central to machine learning is the use of data.

In ML a sufficiently **large dataset** – each data point representing a problem instance – is used to define a set of rules capable of producing some output, representing the solution to the problem instance encoded by the data point. This process is called **training** and uses **optimization techniques** to select the best element, with regard to one or more criteria, from a set of available alternatives.

A concrete example is the definition of an algorithm to classify images of pets: a sufficiently large number of pictures should be available labelled by one among, e.g., two possible alternatives, namely dog and cat.

Using optimization techniques, such as the descent gradient method, training can define a sequence of instructions to classify the subject of a picture as a dog or a cat. Afterwards, the resulting set of instructions is used to classify images that do not belong to the training set, this process is called deployment.

This approach makes it possible to achieve complicated tasks without the necessity to translate vast amounts of data into mathematical/logical functions as with machine-based approaches. Thus, a clear advantage of the modern view of building algorithms is that it does not require a high degree of ingenuity to solve complex problems and instead uses off-the-shelf powerful optimization techniques.

A disadvantage is that the algorithm produced by training may be imprecise, e.g., classifying the subject of a picture as a dog when it is a cat or vice versa, especially when the training data set is not large enough or contains bias. Another disadvantage is that the algorithms produced by training are difficult to understand by humans as they consider too many features to produce a solution. There are approaches to mitigate these disadvantages such as monitoring the performance and adjusting the algorithm if necessary, filtering the dataset used for training to eliminate biases and improve precision, or devising techniques to help humans predict the results produced by algorithms that result from training.

There are different **ML techniques** according to the type of learning and the nature of the input-output relationships within the training data. In **supervised learning**, the human being directly teaches the computer system how to perform certain operations: a set of labelled data (training set) is provided to the system, which contains both input and output data, and from this set of data the algorithm builds a model capable of predicting the response values for a new set of input data. Classification tasks (assigning inputs to predefined categories) and regression tasks (predicting a continuous value) are common applications of supervised learning. **Unsupervised learning** deals with unlabeled data, where the algorithm tries to discover patterns, structures, or relationships within the data without explicit guidance in the form of labelled outputs. Clustering algorithms (grouping similar data points) and dimensionality reduction techniques (simplifying the representation

of data) are common applications of unsupervised learning. Reinforcement learning involves an agent interacting with an environment and learning to make decisions by receiving feedback in the form of rewards or penalties. Applications include game playing (e.g., AlphaGo), robotic control, and autonomous systems.

Deep learning

Deep learning (DL) employs the so-called **neural networks**, i.e., computational models inspired by the composition of biological neural networks and trying to simulate the human brain's decision processes. They consist of interconnected nodes, also known as neurons or units, organised in one or more layers. In case of more layers, each layer is connected only with those immediately below and above it. The three main types of layers are input, hidden, and output layers.

In the **input layers**, neurons receive the initial data or features of the input. Each neuron represents a feature, and the input layer acts as the starting point for information processing.

The **hidden layers** come between the input and output layers. The term 'hidden' arises because here neurons do not directly interact with the external environment or the final output, but they apply weights to the input data and use activation functions to introduce non-linearities to the model, enabling it to learn complex patterns.
In **output layers**, neurons produce the final result or prediction based on the processed information from the hidden layers. The number of neurons in the output layer depends on the nature of the task (e.g., classification or regression).

There are two key aspects common to the various deep learning techniques: (i) the elaboration of models consisting of several levels or stages of non-linear information processing; and (ii) methods for learning representation at progressively higher and more abstract levels. Accordingly, in DL the inner layers of neural networks are not designed by human engineers but are learned by the system autonomously through data collection and processing, employing general learning methods.

Traditional v. Modern AI

Models falling under the "traditional" etiquette are **model-based** systems, which follow rules that were predefined by the developer. "**Traditional AI**" is mainly based on a symbolic approach. Symbolic systems, in contrast to neural networks, rely on explicit rules and representations of knowledge. These systems manipulate symbols and use logical operations to process information. In order to further understand the difference between the two approaches, it is important to underline that symbolic systems follow a top-down approach, where knowledge is introduced by the developer into the knowledge base and the machine translates it into data and symbols. From a set of knowledge, the system deduces certain effects.

Neural networks, instead, follow a bottom-up approach, where the process of constructing a model or system starts with the most basic elements and gradually builds up to more

complex structures. This involves designing models that learn hierarchical representations of data from the raw input. Accordingly, it is assumed that neural networks are capable of more complex representations starting with basic sensory input and progressively forming more abstract and sophisticated concepts, while symbolic systems can solve problems in more limited domains.

As a consequence, traditional models are explainable but often less accurate or suited to accomplish simpler tasks than ML algorithms. For instance, expert systems are a classic example of symbolic systems and use knowledge bases of rules and logical inference engines to make decisions or solve problems in specific domains.

Neural networks, on the contrary, can face changeable, uncertain, incomplete and dynamic situations; at the same time, the peculiar feature of neural networks is their opacity and poor explainability.

Nevertheless, today hybrid approaches are widely used, aiming to combine the symbolic reasoning of traditional AI with the learning and pattern recognition capabilities of neural networks.

Types of AI

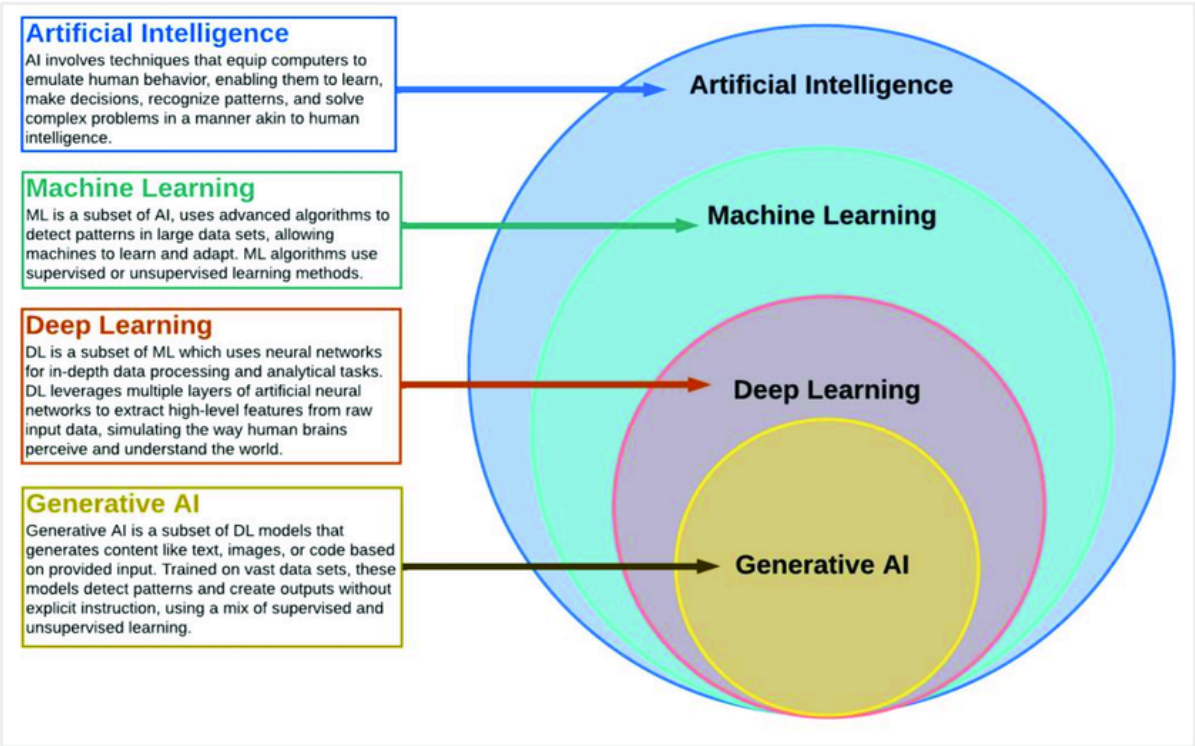


Figure 1: Types of AI. Source: Zhuhadar, Lily Popova & Lytras, Miltiadis, *The Application of AutoML Techniques in Diabetes Diagnosis: Current Approaches, Performance, and Future Directions*, in *Sustainability*, 2023 n. 15, p. 13484.

1.2.2 Defining 'public administration'

The term 'public administration' can also be understood in different ways. In this Handbook, we understand public administration in a broad manner, to include the planning, adoption and implementation of legislation and government policies. To illustrate, some of the areas of public administration covered in the Handbook are:

- **Public service provision**, including both the delivery of services and the provision of information regarding those services. Examples include: water, education, healthcare, housing.
- **Enforcement of laws and government policies**. Examples include: identifying violations of laws and policies by the police; identifying those suspected of violating laws; detection of fraud.
- **Internal management**. Examples include: Recruitment of public sector employees.

1.2.3 Illustrating uses of AI in public administration

AI used in the context of public administration may be designed and developed internally, by a public authority itself, or externally, by the private sector. Before turning to examples of uses of AI in public administration, we will briefly address the question: ***Why* is AI used in public administration?** The answer to this question affects the type and use cases of AI systems in public administration:

- Improve efficiency (e.g. because AI can process information much faster than humans)
- Improve decision-making and policy-making (e.g. higher accuracy/consistency rates may be possible with automated rather than human decision-making)
- Allow civil servants more time to spend on high value work instead of mundane tasks (e.g. simple, repetitive tasks can be done by AI so that civil servants can conduct tasks requiring creative or social intelligence)
- Improve speed and quality of (delivery of) public services[8]

In the following paragraphs, we provide some examples of uses of AI in public administration across Europe, and briefly explain the kind of AI that is typically used for these systems:

[8] The expected benefits of AI in public administration are, for example, listed in several OECD papers, such as OECD (2019), Using digital technologies to improve the design and enforcement of public policies. OECD Digital Economy Papers. February 2019 No. 274, 7.

Public service provision:
AI in public administration can help deliver services and/or automate the transaction of services. In addition, it can help provide information about services.[9] Examples can be found for both kinds.

Delivering services
Finland: Espoo experiment - ultimately not taken further than an original analysis due to ethical concerns, but intended to introduce new, preventive health services.[10]

- *Type of system*: Risk forecasting model
- *How it works*: The City of Espoo in partnership with Tieto ran an experiment to determine if AI could work as a tool to provide services to those who are under risk of experiencing violence or neglect in their homes preventatively. The data set was the combination of customer information from social and health services and early development services from the years 2002-2016.[11] The hypothesis was that AI could detect patterns that would indicate abuse or neglect and thus help social and health care providers take preventative measures. The experiment was seen as a success, but all results and data were destroyed due to their sensitive nature and in anticipation of legislation that would allow for the regular implementation of such programs.[12]

Spain: Misty II robot experiment to see whether robots include quality of life for elderly[13]

- *Type of system*: Social robot/programmable personal robot platform for developers
- *How it works*: The city of Barcelona carried out a pilot test with Misty 2, a robot of the Lleida-based company Saltò Group, in homes for the elderly. The scope of the test was to study the impact the robot could have in improving the quality of life and well-being of the elderly, and their caregivers. Among the others, the impact of helping to reduce the isolation of the elderly, monitoring their health, and improving patient adherence to medication is studied.[14] More specifically, it keeps track of the elderly's health by asking questions about daily habits (sleep quality, medication taking, and

[9] Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government information quarterly*, *39*(3), 101714.
[10] Tietoevry, 'The City of Espoo: A unique experiment with AI' (2018) <https://www.tietoevry.com/en/success-stories/2018/the-city-of-espoo-a-unique-experiment/>.
[11] Espoo City, 'Espoon Kokeilu Todisti: Tekoäly Tunnistaa Tukea Tarvitsevia' (7 June 2018) <https://www.sttinfo.fi/tiedote/68401520/espoon-kokeilu-todisti-tekoaly-tunnistaa-tukea-tarvitsevia?publisherId=3385>.
[12] Salla Vuolteenaho, '"Voi hyvänen aika", huudahtaa virkamies - Tekoäly löysi 280 tekijää, jotka ennakoivat lastensuojelun asiakkuutta' (25 September 2018) <https://yle.fi/a/3-10413353>.
[13] Saltó Group, Barcelona City Council, 'Misty II, a robot for improving the quality of life of the elderly living alone' (8 February 2020) <https://isocial.cat/en/misty-ii-a-robot-to-improve-the-quality-of-life-of-the-elderly-living-alone/>.
[14] Saltó Group, Barcelona City Council, 'Misty II, a robot for improving the quality of life of the elderly living alone' (8 February 2020) <https://isocial.cat/en/misty-ii-a-robot-to-improve-the-quality-of-life-of-the-elderly-living-alone/>.

diet). Additionally, it also helps with the elderly's loneliness as it provides companionship through dancing, telling jokes, and playing music.[15]

Belgium: Automated decision-making system used to assign students to schools to avoid unfair competition with a 'first come first served' system. Students assigned based on their preferences and to consider the distribution of disadvantaged/privileged students.[16]

- *Type of system*: Algorithm
- *How it works*: In Belgium, schools used the rule 'first come, first served' for enrollment, causing parents to camp in tents in front of schools to be the first ones to enrol their children. This was considered unfair; therefore, in 2018, the Flemish government introduced a central online system that uses an algorithm that decides the school where to enrol each student. Once assigned, each student receives a ticket and can enrol in the assigned school by using the ticket. This system is considered fair because the algorithm assigns as many students as possible to a school of their highest possible preference, and the list of students who cannot enrol in the schools of their choice is minimised.[17]

<u>Providing information about services</u>

Various chat bots answering questions about changes of regulation, taxes, subsidies, visa applications etc. exist or are being developed. For example, Sofie in Norway, a chatbot used by the Norwegian Tax Administration to assist with tax-related queries, providing information on filing taxes and other related services or those introduced during COVID-19.[18]

- *Type of system*:: AI-powered computer chatbot program
- *How it works*: Example Advokat (provided by a non-profit organisation): The system was born in California during COVID-19. Since in the US, there is no national, universal, equitable, paid leave it was not simple, during the health crisis, for workers to know their rights regarding leave and pay. The system, called Advokat, explains to workers their rights in regard to protected leave and pay in a personalised, accurate, and easy way. The responses are based on workers' workplace, tenure, and medical needs.

Enforcement of laws and government policies:

Hungary, Greece

- *Type of system: Face recognition system*

---

[15] Cristina Getson and Goldie Nejat, 'Socially Assistive Robots Helping Older Adults through the Pandemic and Life after COVID-19' (2021)10 106.

[16] Automating Society Report 2020 (Algorithm Watch) <https://automatingsociety.algorithmwatch.org/>.

[17] Koen Vervloesem, 'In Flanders, an algorithm attempts to make school choice fairer' (Algorithm Watch) <https://automatingsociety.algorithmwatch.org/report2020/belgium/belgium-story/>.

[18] 'New Chatbot Explains Workers' Rights to Leave and Pay During COVID-19' <https://legalaidatwork.org/new-chatbot/>.

- A pilot project by iBorderCtrl, funded by Horizon2020 with 4 million Euros, introduced AI-powered lie detectors at airport border checkpoints in Hungary and Greece. The system monitors people's faces for signs of lying, and those flagged undergo further screening by a human officer. Critics, including EDRi and Amnesty International, raised concerns about fundamental rights abuses, including discrimination and data protection. Greek NGO Homo Digitalis filed a petition to the Greek Parliament, highlighting transparency issues, and MEP Patrick Breyer challenged the EU Commission's refusal to disclose ethical assessments of the iBorderCTRL system.[19]

Netherlands and Belgium: Use of cameras using AI to see whether people driving are holding their mobile phones.[20]

- *Type of system*:: Algorithmic Camera system/Deep learning system
- *How it works*: The police in both the Netherlands and Belgium have implemented smart traffic cameras that are able to detect people using their phones while driving. In the Netherlands, the AI is able to read and decode the image taken by the Monocam camera in order to detect the violation and automatically send it to the CJIB (public authority responsible for traffic fines).[21] Yet, this system is not without its errors, as there have been instances where the AI has misread the image and fines were given out to people who were not in violation of any traffic laws.[22]

Belgium: Use of facial recognition technology by police to help with investigations e.g. finding suspects and missing children, but also to send alerts to the police for violations of rules e.g. trucks driving outside school during certain hours when they should not be due to the high number of children in the area.[23]

- *Type of system*: Facial recognition technology/Deep learning system

---

[19] https://edri.org/our-work/immigration-iris-scanning-and-iborderctrl/, https://www.homodigitalis.gr/wp-content/uploads/2018/11/05.11_HomoDigitalis_Petition_iBorderCtrl.pdf

[20] Openbaar Ministerie, Openbaar Ministerie start digitale handhaving op handheld telefoongebruik achter het stuur' (12 November 2020) (Public Prosecution Service, Public Prosecution Service starts digital enforcement against the use of handheld phones behind the wheel). <https://www.om.nl/actueel/nieuws/2020/11/12/openbaar-ministerie-start-digitale-handhaving-op-handheld-telefoongebruik-achter-het-stuur> (NL) and Vias Institute, 'Succesvolle test met camerasysteem om gsm-gebruik achter het stuur te detecteren' (15 December 2020) <https://www.vias.be/nl/newsroom/succesvolle-test-met-camerasysteem-om-gsm-gebruik-achter-het-stuur-te-detecteren-/>. (Belgium), both cited in Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government information quarterly*, *39*(3), 101714.

[21] Isabelle Fest, 'Values? Camera? Action! An ethnography of an AI camera system used by the Netherlands Police' (Research Square, 13 June 2023).

[22] Jerusalem Post staff, 'AI falsely accuses, fines artificial intelligence expert of using phone while driving - report' (18 February 2024) <https://www.jpost.com/business-and-innovation/tech-and-start-ups/article-787534>.

[23] Pieter Haeck, 'Belgian police watchdog rules use of Clearview AI "unlawful"' (3 October 2022) <https://subscriber.politicopro.com/article/2022/03/belgian-police-watchdog-rules-use-of-clearview-ai-unlawful-00016045>.

- *How it works*: The Belgian Police made use of a facial recognition system developed and operated by a US company Clearview AI. The database for comparison was collected from publicly available sources such as social media, news, mugshot websites and other open sources.[24] Since the database was gathered and operated by Clearview, the gathered data was not restricted to any particular jurisdiction and was mainly used for international investigations.[25] At the time Belgian law did not permit the use of facial recognition technology and raised serious concerns about the privacy implications of the actions of the police in this instance.

Netherlands: Use of risk-profiling systems to detect persons considered to be at high risk of committing (welfare) fraud.

Tax Authority

- *Type of system:* Machine learning algorithm
- *How it works:* The tax Authority used an automated risk selection system to filter out cases with a high risk of errors and fraud amongst Child Care Benefits applicants. The risk classification model was trained over time based on regular and irregular benefits applications. The more an application resembled an irregular application, the higher the assigned risk score. Indicators used in the risk classification system included nationality and income. Applications with a high risk score would be further investigated for fraud.[26]

Dienst Uitvoering Onderwijs (Provision of conditional student grant)

- *Type System:* Rule Based Algorithm
- *How it Works:* DUO made use of an algorithm to calculate the risk of fraud amongst students receiving conditional student grants. The risk was calculated based on objective variables, namely, age, type of education, and proximity in registered address between student and parent. The risk profile was created based on workshops with various governmental parties on risk profiling.[27] Students with the highest risk scores were placed on a selection list and had a higher chance of being selected for further investigation.[28]

---

[24] Knowledge Centre Data and Society, 'Belgium - Unlawful use of facial recognition by federal police' (February 2022) <https://data-en-maatschappij.ai/en/policy-monitor/belgium-unlawful-use-of-facial-recognition-by-federal-police>.

[25] Nick Amies, 'Clearview facial recognition technology raises severe privacy concerns' (15 October 2023) <https://www.brusselstimes.com/731979/clearview-facial-recognition-technology-raises-severe-privacy-concerns>.

[26] Autoriteit Persoonsgegevens, 'Belastingdienst/Toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag' (2018) <https://open.overheid.nl/documenten/ronl-71b93874-d34e-488a-9caf-9c455dec34a0/pdf> pg 14-18.

[27] Intern controle Uitwonende beurs 18-19

[28] Auditdienst Rijk, 'Onderzoeksrapport DUO Beheersing privacyaspecten proces uitwonendencontrole' (14 September 2023) <https://open.overheid.nl/documenten/dpc-7f8bb1dd56a49bff50b7958d48aba0b1909f3c3c/pdf>.

Municipality of Rotterdam

- *Type of system:* Machine Learning System
- *How it works:* Using previous examples of welfare fraud, Accenture and the municipality of Rotterdam trained an algorithm to calculate the risk of fraud among welfare recipients.[29] A person's risk score was calculated based on 315 objective and subjective variables, including neighbourhood, age, gender, Dutch speaking and writing skills, debts, and psychological and physical problems.[30] Benefit recipients with the highest risk scores were subjected to additional supervision by enforcement agents.[31]

City of Amsterdam

- Type of system: Crime Anticipation System
- The Crime Anticipation System (CAS) in Amsterdam is a predictive policing tool that aims to forecast specific crimes in designated areas within a two-week period. It utilizes machine learning to analyze socio-economic data, historical crime data, and geospatial information to categorize areas as red, orange, or yellow, indicating varying levels of risk. However, CAS's indicators, including the number of 'non-Western allochtones' in an area, raise concerns about ethnicity-based profiling. By singling out specific ethnicities, CAS incorporates questionable societal categorizations and biases, posing data protection risks and fostering discrimination based on ethnic origin. While not directly targeting individuals, place-based predictive policing systems like CAS may contribute to overpolicing certain communities, reinforcing existing biases and inequalities in policing data.[32]

Internal management

Sweden: AI in Tengai robot used in the Swedish municipality Upplands-Bro to make public administration recruitment processes less biassed.[33]

- *Type of system*: Machine Learning System
- *How it works*: Tengai has developed an interview robot in attempts to cut down on the bias that exists within in person interviews. The robot was developed to

[29] Reinier Tromp, 'Fraude-algoritme op de snijtafel' (6 March 2023) <https://www.vpro.nl/argos/lees/onderwerpen/algoritme/2023/fraude-algoritme-op-de-snijtafel-rotterdam-verdenkt-vooral-jonge-moeders.html>.

[30] Eva Constantaras and others, 'Inside the Suspicion Machine' (6 March 2023) <https://www.wired.com/story/welfare-state-algorithms/>; Reinier Tromp, Rik Delhaas and David Davidson ' In het vizier van het Rotterdamse bijstandsfraude-algoritme?' (18 December 2021) <https://www.vpro.nl/argos/lees/onderwerpen/artikelen/2021/in-het-vizier-van-het-rotterdamse-algoritme-.html>.

[31] Eva Constantaras and others, 'Inside the Suspicion Machine' (6 March 2023) <https://www.wired.com/story/welfare-state-algorithms/>; https://www.ftm.nl/artikelen/algoritme-gemeente-rotterdam.

[32] https://ceur-ws.org/Vol-2103/paper_6.pdf

[33] Henry Bewicke, 'Sweden is using robots to conduct its job interviews in a fairer way' (18 July 2019) <https://www.weforum.org/agenda/2019/07/sweden-robot-remove-bias-from-recruitment/>.

achieve goals such as collecting objective data from interviews, achieve a fair candidate experience, increase efficiency, improve the quality of hiring and ensure security and privacy throughout the recruitment process.[34] The robot was trained not to take into consideration factors such as appearance, accent, age or any other factor that might affect the opinion of the interviewer. Further, the robot is made as life-like as possible whereby the robot smiles, nods along and gives human feedback during the interview. The municipality of Upplands-Bro has implemented the Tengai robot in their recruitment process in order to fight discrimination in hiring in the public sector.[35]

United Kingdom: Early Help Profiling System (EHPS) which was used in London, United Kingdom, to help public managers analyse a large variety of data to identify children and families which are vulnerable or at risk of child abuse. Discontinued after ethical concerns.

- *Type of system*: Advanced analytics/advanced modelling techniques/predictive system
- *How it works*: the EHPS was developed by Xantura and collects information from various sources and uses them, through predictive analytics, to identify the most vulnerable children and families. The data collected are from different government agencies (school attendance, housing information, economic indicators etc.). The system favours early intervention without waiting for a crisis to occur. Every month the social workers receive a list of 20 'high-risk' families not yet known to the social workers, and an assessment is made on whether to conduct an early intervention. The system is considered to have a high accuracy as more than 80% of the alerts are valid. The system was designed to save around one million euros per year thanks to the early interventions. However, it was discontinued due to ethical concerns such as the nature of the data collection and the risk of assessment.[36]

More examples from the EU could be added to this short overview and many more will be addressed in this handbook. What this overview shows is that the use of AI in public administration has the potential to significantly transform various aspects of public administration, including public service delivery, enforcement of laws and government policies, and the internal management of the administration itself. AI technologies are meant to streamline public services by automating routine tasks, improving efficiency, and facilitating interactions with citizens. In enforcement, AI is meant to aid in analysing vast data sets to detect anomalies, and ensure compliance with regulations. Internally, AI is intended to optimise resource allocation, enhance decision-making processes, and facilitate transparent management. However, the implementation of AI must be carefully managed to uphold

---

[34] Tengai, 'Robot Interview: Everything a recruiters need to know' (2024) <https://tengai.io/resources/tengai-robot/>.

[35] Henry Bewicke, 'Sweden is using robots to conduct its job interviews in a fairer way' (18 July 2019) <https://www.weforum.org/agenda/2019/07/sweden-robot-remove-bias-from-recruitment/>.

[36] Ed Sheridan, 'Town Hall drops pilot programme profiling families without their knowledge' (30 October 2019) <https://www.hackneycitizen.co.uk/2019/10/30/town-hall-drops-pilot-programme-profiling-families-without-their-knowledge/>.

fundamental rights and ensure non-discrimination. This can be done at different stages of the algorithmic life cycle around which this Handbook is structured.

1.3 Structure of this Handbook

The algorithmic lifecycle is a framework used to understand the emergence and prevention of outcomes clashing with fundamental rights. Several variations of the algorithmic lifecycle exist.[37] However, three general stages can be discerned: the design of the algorithmic system, its organisational embedding, and its use in decision-making. Scholars from differing disciplines have increasingly demonstrated that outcomes clashing with fundamental rights stem from unmitigated risks throughout the lifecycle stages of algorithmic systems.[38] This observation is likewise reflected in legal and ethical instruments used in practice to identify and mitigate such risks and outcomes.[39]

I. Algorithmic design

The algorithmic design refers to how the algorithmic system is built by the organisation. In this stage, it is important to analyse the organisational decisions about training data, the complexity of the algorithmic system, predictive indicators, labels, accuracy, and transparency can all contribute to outcomes clashing with fundamental rights.[40] Examples include the use of biassed training data, the inclusion of discriminatory predictive indicators, and the use of opaque 'black box' systems that make it difficult to identify discrimination and bias.[41]

---

[37] Marabelli, M., S. Newell & S. Handunge, 'The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges', The Journal of Strategic Information Systems 2021-30(3); McGregor, L., D. Murray & V. Ng, 'International human rights law as a framework for algorithmic accountability', International & Comparative Law Quarterly 2019-68(2); Yeung, K., A. Howes & G. Pogrebna, 'AI Governance by human rights-centred design, deliberation and oversight: An end to ethics washing', in: M. Dubber, F. Pasquale & S. Das (eds.), The Oxford Handbook of Ethics of AI. Oxford University Press 2020; Tilburg University, Non-Discrimination by Design. Tilburg University 2021; De Silva, D. & A. Damminda, 'An artificial intelligence life cycle: From conception to production', Patterns 2022-3(6).; Azzutti, A., W. Ringe & A. Stiehl, 'Regulating AI trading from an AI life cycle perspective', in: N. Remolina & A. Gurrea-Martinez (eds.), Artificial Intelligence in Finance: Challenges, Opportunities and Regulatory Developments, Edward Elgar 2023.

[38] Ibid.; Haitsma L.M., The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy, Recht der Werkelijkheid, 2023 (44)2.

[39] Regulation 2016/679 on General Data Protection (GDPR), Art. 35; Gerards, J., M.T. Schafer, A. Vankan & I. Muis, Impact Assessment Fundamental Rights and Algorithms (FRAIA). Ministry of Interior and Kingdom Relations 2022; Utrecht University Data School, Data Ethics Decision Aid (DEDA). Utrecht University 2022; Tilburg University, Non-Discrimination by Design. Tilburg University 2021; High Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI. European Commission 2019.

[40] Marabelli, M., S. Newell & S. Handunge, 2021.

[41] Marabelli, M., S. Newell & S. Handunge, 2021; Fountain, J., 'The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms', Government Information Quarterly 2022-39; Xenidis, R. & L. Senden, 'EU Non-discrimination law in the Era of artificial intelligence: Mapping the challenges of Algorithmic discrimination', in: U. Bernitz, X. Groussot, J. Paju, S.A. & De Vries (eds.), General Principles of EU Law and the EU Digital Order. Kluwer Law International 2020.

II. Organisational embedding

Organisational embedding refers to how the algorithmic system is embedded and configured within the working process of an organisation and how employees interact with the system.[42] Here, organisations must implement meaningful human-in-the-loop and human oversight safeguards to ensure that outputs of (semi-)automated systems are understandable, contestable, and explainable. Despite the necessity of these safeguards, introducing a human element into (semi-)automated processes can bring a host of risks that aggravate negative algorithmic consequences.[43] Such risks include rubber-stamping algorithmic outputs, automation bias, algorithmic illiteracy, boredom and fatigue, and noise.[44] As such, organisations should have procedures and protocols in place to safeguard the quality of decisions taken.[45]

III. Decision-making based on algorithmic output and manual review

The last stage refers to the use of the algorithmic system and its output in decision-making following the manual review.[46] Here, unmitigated design flaws and poor organisational embedding can influence the quality of decisions taken.[47] For example, using an opaque and discriminatory algorithmic profiling system without meaningful and effective human intervention can lead to prolonged discriminatory decisions.[48] For this reason, monitoring and evaluating decisions taken based on the algorithmic system and manual review, and the effectiveness of remedy mechanisms is paramount.[49] This helps to ensure that outcomes clashing with fundamental rights are timely identified and that such outcomes are not fed back into the algorithmic systems and organisation as new data.

---

[42] Marabelli, Newell & Handunge 2021, pp. 4-6.

[43] Koulu, R. (2020). Proceduralizing control and discretion: Human oversight in artificial intelligence policy. *Maastricht Journal of European and Comparative Law*, *27*(6), 720-735.

[44] Binns, R. (2022). Human Judgement in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, *16*(1), 197-211.; Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388-409.; Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: a flaw in human judgement*. Hachette UK.

[45] Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity*, *23*(1), 3-28.

[46] Marabelli, Newell & Handunge 2021, pp. 2-3.

[47] Haitsma L.M., The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy, Recht der Werkelijkheid, 2023 (44)2.

[48] Marabelli, Newell & Handunge 2021, pp. 6-7; McGregor, Murray & Ng 2019, pp. 323-324.

[49] Marabelli, Newell & Handunge 2021, pp. 6-7; Yeung, Khan, Kalra & Osoba 2021, p. 20; Tilburg University 2021, pp. 61-68.

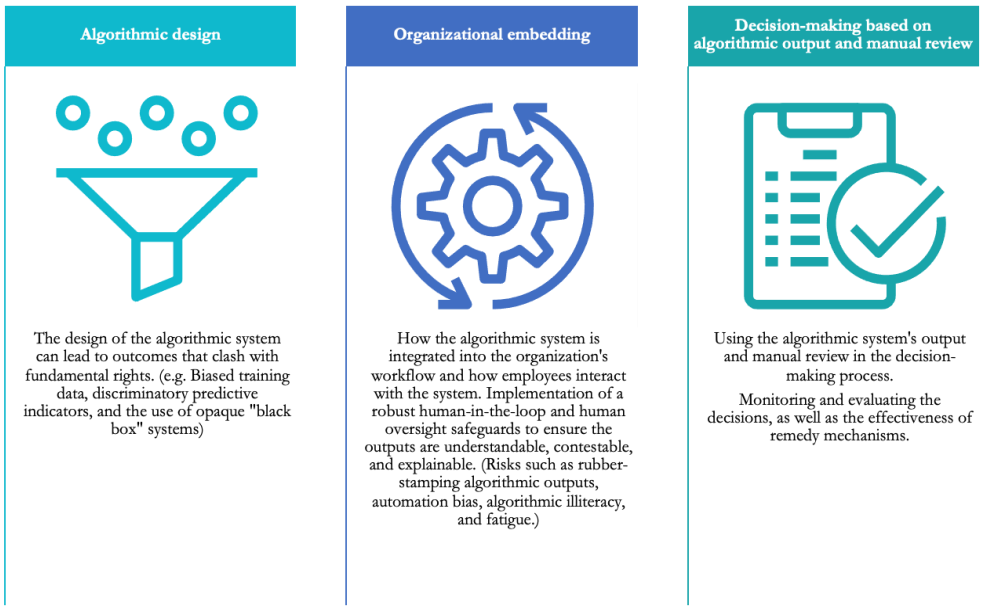


Figure 2. The algorithmic lifecycle.

1.4 Methodology[50]

This Handbook does not present an exhaustive overview of the use of AI in public administration in the European Union nor of the legal acts regulating its use. Rather, it explores possible threats to fundamental rights, especially to the right to non-discrimination, at the different stages of the algorithmic life cycle and discusses possible solutions. It provides guidance to the participants of the training events as to how they can approach cases by explaining how such AI systems work and are used, by identifying possible risks of AI systems to fundamental rights, and by sketching the applicable (emerging) law and policy frameworks. It does not provide definite answers but should be seen as a source of inspiration for judges, lawyers and other legal experts.

This Handbook is part of the JuLIA project which investigates different areas in which AI (potentially) impacts fundamental rights - judicial decision making, health, markets, and data protection. The Handbooks that are accompanying this project are interdisciplinary in nature, practice oriented, with a focus on the EU legal framework and its interconnectedness with national developments. By focussing on practices and general principles, the Handbooks provide a guide to legal practitioners to assess these practices in light of a more effective protection of fundamental rights. The Handbooks complement the training events and workshops, another central feature of the JuLIA project. They form a basis of discussion for the participants of these workshops. These are designed to benefit both judges with experience with AI cases as well as newcomers to the field. At the workshops judges and experts are discussing the possible impact of AI tools in the respective fields. In practical small group sessions, they can exchange practices and ideas using hypothetical cases, which

[50] This section is based on the methodology sections of the other JuLIA Handbooks written by P. Iamiceli and M. Artigot i Golobardes.

are based on real cases, if available. In the field of AI in public administration, there are a number of mainly national cases that can serve as an inspiration, many have been included in this case book.

As a tool for active training, this Handbook is meant to grow along the training experience (and help develop an e-learning module). It will benefit from the contributions of judges interested in participating in training events and the contributions developed during the training events and after the training will be integrated. As in previous judicial training projects (see Re-Jus, FRICoRe), mutual learning is the key. Based on this collaboration, this set of materials (still in draft format and not – yet - due to circulate outside the group of workshop participants) will be integrated in the forthcoming months with special regard to the presentation of experimental practices. Users of this Handbook, both as trainers and trainees, are warmly invited to share their feedback and possible contributions with the Project leaders.

2. Algorithmic Design

Lucas Haitsma, Linda Ardenghi, Beatrice Ancuceanu, Ida Varošanec, Lorena Flórez

2.1 Risks for fundamental rights in the design stage of AI systems used in public administration

During the decision-making process of any public entity considering buying or designing their own AI system, it is crucial to identify and thoroughly understand the problem they aim to solve.[51] Design thinking, as a problem-solving framework, offers the possibility of improving products, processes, and services through innovation-based practices.[52] One of the critical points in the design thinking methodology is ideation, where the entity should establish which problem needs to be addressed and decide if technology, specifically AI, is indeed the solution for this problem. It is important to recognize that AI is not always the answer to every challenge. The decision to embark on AI projects should be taken consciously, evaluating the pros and cons of each potential solution. This stage is often one of the longest in the process because many entities, in their rush to join the AI revolution, claim to need the technology when they have not yet fully comprehended the problem they want to tackle. Taking the time to properly identify and analyse the issue at hand can lead to more effective and appropriate solutions, whether they involve AI or other approaches (*ideation*).

During the design stage, an organisation builds an algorithmic system. This system will later be embedded in an organisation and potentially used in decision-making processes. Within the design stage, decisions are made about, for example, which data to collect, what type of algorithm is used to process the data, and which variables and indicators are relevant to the processing (*data sets and training data*).[53] Additionally, an organisation will need to decide whether to build the algorithmic systems in-house or build on top of a system developed elsewhere, as is often the case (*third-parties*). In this stage, design flaws include collecting protected categories of data, using biassed training data, using discriminatory predictive indicators, and making use of opaque algorithmic systems.[54] If such design flaws are not

---

[51] Flórez Rojas, M.L. and Leal Vargas, J. (2020). The Impact of Artificial Intelligence Tools: An Analysis of the Public Sector in Colombia in Carolina Aguerre (ed), Artificial Intelligence in Latin America and the Caribbean: Ethics, Governance and Policies, CETyS, FairLAC, 23-25. <https://proyectoguia.lat/wp-content/uploads/2020/10/7.-GECTI-eng.pdf>. See also Berryhill, J., et al. (2019), Hello, World: Artificial intelligence and its use in the public sector, OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, 115-117. <https://doi.org/10.1787/726fd39d-en.>

[52] Mayer et. al. (2021. Assessing the Impact of Design Thinking in Organizations: Foundations of a Framework, *Design Thinking Research*, *255*, 23-26.

[53] Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, *30*(3), 101683. P 2-4 in Haitsma, L. (2023). The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy. *Recht der Werkelijkheid*, *44*(2), 61-83.

[54] Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, *30*(3), 101683. P 2-4. Fountain, J. E. (2022). The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms. *Government Information Quarterly*, *39*(2), 101645. Pp. 3-4; Xenidis, R., & Senden, L.

identified and mitigated, they can contribute to outcomes clashing with fundamental rights.[55] For example, an overrepresentation of certain groups stemming from biassed training data and discriminatory selection criteria can clash with the right to equal treatment and non-discrimination. Additionally, an inability for public administration organisations to understand and explain their decisions due to the use of an opaque algorithmic system can clash with principles of good administration and the right to an effective remedy.

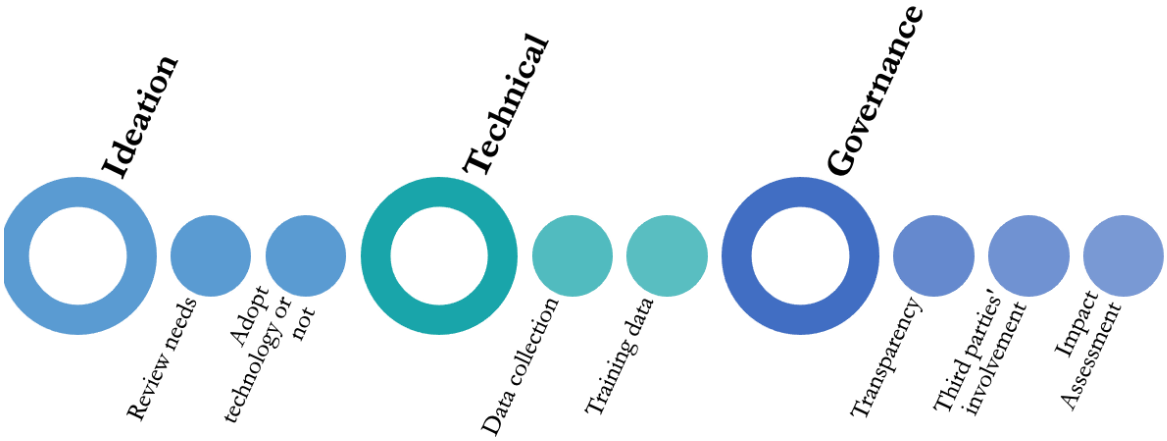


Figure 3. Elements of the algorithmic design of AI systems stage. By Lorena Flórez.

2.2 European legal frameworks and case law

As stated in the introduction to this Handbook, the legal framework on AI and fundamental rights is still developing. This is the case at the EU level, but also at the international and national levels. In the following paragraphs, we introduce key legal provisions on algorithmic design and fundamental rights applicable to the context of public administration. Most of the EU regulations concerning AI are transversal, meaning they apply across all stages of the AI life cycle. However, within each stage—such as design, embedding and decision-making—specific provisions address the unique challenges and requirements of that stage. Additionally, some cases may cross multiple stages due to their significance and the depth of analysis required, serving as key examples that illustrate different aspects of the AI life cycle.

For the EU these are the Charter of Fundamental Rights of the EU, the General Data Protection Regulation, the Law Enforcement Directive and the Artificial Intelligence Act. For the Council of Europe these are the European Convention on Human Rights and the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. We also introduce relevant case law from the Court of Justice of the European Union as well as national courts.

(2020). EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination in Ulf Bernitz et al (eds), General Principles of EU law and the EU Digital Order. 154-159.

[55] Haitsma, L. (2023). The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy. *Recht der Werkelijkheid*, *44*(2), 61-83.

2.2.1 The European Union

Charter of Fundamental Rights of the European Union

The Charter guarantees the fundamental rights that individuals enjoy within the EU. Key articles to consider in relation to algorithmic design include: Article 7 on the right to respect for private and family life; Article 8 on the right to protection of personal data; Article 20 on the right to equal treatment; Article 21 on the right of non-discrimination; and Article 52(1) on limitations to fundamental rights.[56]

General Data Protection Regulation (GDPR)

The General Data Protection Regulation has been in effect since May 2018 and aims to harmonise data protection laws across the European Union. In the context of algorithmic design, compliance with the GDPR is essential, particularly regarding the collection, storage, and processing of training data, to safeguard individuals' rights to privacy and data protection. Key provisions relevant to the design stage include; article 5 on the principle relating to the processing of personal data; Article 6 on the lawfulness of processing; Article 9 placing limits on the processing of special categories of data; and article 25 obliging controllers to implement technical and organisational measures designed to implement data protection principles.[57]

Law Enforcement Directive (LED)

The Law Enforcement Directive, effective since May 2018, regulates the use of personal data by law enforcement agencies within the European Union. This Directive establishes regulations for protecting individuals' personal data processed by law enforcement for preventing, investigating, detecting, or prosecuting criminal offences, as well as for executing criminal penalties. This legal instrument contains provisions similar to the GDPR, specifically tailored to a law enforcement context.[58] In terms of design requirements, LED requires technical measures designed to implement data protection principles.[59]

---

[56] European Union, *Charter of Fundamental Rights of the European Union*, 2012/C 326/02, 26 October 2012

[57] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

[58] Directive 2016/680 (2016). 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 2008/977/JHA (O.J. 2016., L119.)

[59] See article 29 LED.

Artificial Intelligence Act

The Artificial Intelligence Act aims to establish unified regulations governing the development and utilisation of AI within the European Union. By setting requirements for transparency, accountability, and human oversight, the AI Act contains key provisions regulating algorithmic systems and their design.[60] In particular, with regard to high risk systems delineated in Annex III of the regulation, provisions regulating the design include; article 9 requiring a risk management system throughout the lifecycle of high risk systems; article 10 regulating data and data governance; article 14 requiring the possibility of an effective oversight; article 15 pertaining to the accuracy of AI systems; and article 17 requiring providers of AI systems to ensure a quality management system.[61]

2.2.2. Relevant case law

Automated processing of Passenger Name Record (PNR)

Through issuing Opinion 1/15 of July 2017 the CJEU prevented the conclusion of a PNR agreement between the EU and Canada by identifying several weaknesses in the draft with relation to articles 7 and 8 CFEU, thus arguing that the agreement seems incompatible with those articles.[62] The Court found that an objective interest of the EU justified the interference and as such did not compromise the essence of article 7 and 8 CFEU.[63]

In relation to the automated processing of PNR data of passengers flying to Canada, the CJEU noted that such processing of PNR data and the use of automated analysis against predetermined models and criteria presented margins of error that were found to be significant by the European Data Protection Supervisor (EDPS).[64] Such pre-established models and criteria must be specific, up to date, non-discriminatory and reliable so as to target individuals actually suspected of committing an offence falling within the scope of the agreement.[65] Regarding the retention of data, the CJEU noted that retaining the data of all passengers flying in and out of Canada, even after it has been established that they pose no direct or indirect risk to national security, is not sufficiently limited to what is strictly necessary and proportional and thus is incompatible with articles 7 and 8 CFEU.[66]

---

[60] 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 (Artificial Intelligence Act) [2024] OJ L2024/1689 (AI Act).
[61] Ibid. art 9, 10, 15, 17.
[62] Opinion 1/15 of the Court of Justice of the European Union of 26 July 2017 on the draft PNR agreement between Canada and the European Union, EU:C:2017:592.
[63] Ibid., para 151.
[64] Ibid., para 169-170.
[65] Ibid., para 172.
[66] Ibid., para 204-206.

Ligue des droits humains

On 21 June 2022, the CJEU delivered a landmark judgement in the *Ligue des Droits Humains* case concerning discrimination and algorithmic profiling in the context of border security.[67] It interpreted whether the PNR Directive, which provides the legal basis for profiling, based on Passenger Name Records data (hereafter PNR data), is compatible with the fundamental right to privacy and data protection, and the right to non-discrimination.[68] Under the PNR Directive, air carriers must transfer passenger data to passenger information units within member states, where it is automatically processed against law enforcement databases and predefined risk indicators to identify and select passengers posing a threat of serious crime or terrorism[69]

The CJEU firstly stated that data collected under the PNR Directive must be limited to what is necessary to combat serious crime and terrorism, and must exclude sensitive data.[70] To this extent, open data categories enabling the inclusion of information that could reveal sensitive information, such as the category of 'general remarks' were restricted through interpretation.[71] Secondly, the judges in Luxembourg stated that when assessing PNR data against 'relevant databases', only those explicitly listed in the Directive could be used. Furthermore, assessing PNR data against such databases must comply with the right to non-discrimination, and entry into these databases must be based on objective and non-discriminatory factors.[72]

Thirdly, concerning the predetermined criteria that PNR data could be assessed against, the CJEU stated that criteria must 'in no circumstances' be directly or indirectly discriminatory. Therefore, the criteria must not only be non-discriminatory in their wording but also must not disadvantage individuals from a protected group. Furthermore, the Court prohibited the use of self-learning artificial intelligence technologies that could modify assessment processes and criteria, and how those criteria were weighted, without any intervention or review by a human. Here the court stated that the use of such systems would likely hinder the ability to review positive matches and understand why passengers were selected for additional supervision, thereby hindering data subjects ability to challenge the non-discriminatory nature of the results of the system.[73]

---

[67] The facts of this case within this book are based on a previous summary made by one of the contributing authors in Haitsma, L. M. (2023). Regulating algorithmic discrimination through adjudication: the Court of Justice of the European Union on discrimination in algorithmic profiling based on PNR data. *Frontiers in Political Science*, *5*, 1232601.

[68] CJEU, Judgment of 21 June 2022, Ligue des droits humains ASBL v Conseil des ministers (C-817/19), ECLI:EU:C:2022:491 ('*Ligue des droits humains*').

[69] Haitsma, L. M. (2023). Regulating algorithmic discrimination through adjudication: the Court of Justice of the European Union on discrimination in algorithmic profiling based on PNR data. *Frontiers in Political Science*, *5*, 1232601.

[70] *Ligue des droits humains,* paras 117, 128, 130

[71] Ibid. paras 128, 136.

[72] Ibid. paras 187 to 191.

[73] Ibid. paras 194 to 197.

2.2.2 The Council of Europe

European Convention on Human Rights (ECHR)

Over time, the ECtHR has established the connection between the ECHR provisions drafted in 1950's and AI in its jurisprudence, which will be addressed in more detail in the next section. In particular, case law will address Article 8 which enshrines the right to respect for private and family life, and Article 14 which prohibits discrimination on grounds such as sex, race, religion, and other status.[74]

Protocol No 12 to the ECHR, which came into force in 2005, bolsters protections against discrimination by extending the scope of Article 14 to encompass discrimination in any right set forth by law. This protocol significantly strengthens the ECHR's anti-discrimination measures, reinforcing the commitment to equality and non-discrimination across various spheres of life.[75]

Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108)

This protocol aims to modernise data protection standards in the digital age.[76] It expands Convention 108's scope to cover processing of personal data, ensuring comprehensive protection against potential risks associated with data processing activities.[77] It emphasises principles such as transparency, fairness, and accountability in data processing practices, bolstering individuals' privacy rights in an era marked by rapid technological advancements.[78] These principles also play an important role at the design stage.

2.3 National case law and legal frameworks: Examples from Europe and beyond

2.3.1 Legislation

Many aspects of existing national legislation will be applicable to the matters of algorithmic design. Consequently, there are not many legislative changes specifically pertaining to AI development. This is understandable considering the existing EU regulations governing the different design aspects and the introduction of the AI Act which will regulate these matters. Consequently, what we can see on the national level is either adaptations of extant laws or the inclusion of AI-related provisions within. For instance, France has incorporated AI-related

[74] Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, Nov. 4, 1950, ETS 5.

[75] Council of Europe, Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 2000, ETS 177

[76] Council of Europe, Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data: convention 108+ (explanatory notes) https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/LIBE/DV/2018/09-10/Convention_108_EN.pdf

[77] Ibid.

[78] Ibid.

provisions within its broader digital and innovation laws, particularly concerning the use of AI in sensitive areas like healthcare and public administration.[79]

2.3.2 National case law

United Kingdom: R v South Wales

The case revolves around the legality of South Wales Police Force's (SWP) use of live automated facial recognition technology (AFR) in an ongoing trial with a system called AFR Locate. This system employs surveillance cameras to capture digital images of the public, which are then processed and compared with images of individuals on a law-enforcement watch list. The system was used in an overt manner. The complaint lodged by the applicant was based on the grounds that AFR is incompatible with the right to respect for private life under Article 8 ECHR.[80]

AFR assesses whether two facial images are of the same person by taking a digital photograph of a person's face and processing it to extract biometric data (i.e., measurements of facial features).[81] This data is then compared with facial biometric data from images in a law-enforcement watch list.[82] When AFR Locate is deployed, SWP mounts CCTV cameras on police vehicles or on poles/posts to capture images of the faces of anyone passing within the camera's range.[83]

In the first round of proceedings, the Divisional Court found that the interference with Article 8(1) was justified by the conditions in Article 8(2) ECHR.[84] The first ground of appeal then questioned whether the Divisional Court erred in concluding that the interference was in accordance with the law for the purpose of Article 8(2) ECHR.[85]

The main point of interest of this case within the algorithmic design stage is the issue of indirect discrimination. The Court examined the Public Sector Equality Duty (PSED) outlined in section 149(1) of the Equality Act, which mandates that UK authorities eliminate discriminatory practices and promote equal treatment for all individuals. The Court noted that this duty covers specially protected categories such as sex and race. The applicant argued that SWP breached PSED as they never had due regard to the need to eliminate discrimination on the grounds of race and sex, which may arise from the software used in the deployment of AFR Locate. The Court acknowledged the existence of scientific evidence demonstrating that some forms of facial recognition softwares can be biassed and create a greater risk of false identification of people of colour, asian, other minority ethnic

[79] See for instance Article L311-3-1 of the French Public Relations and Administration Code which requires authorities to mention when an algorithm has been used to make a decision.
[80] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 2.
[81] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 8-9.
[82] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 8.
[83] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 12.
[84] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 37.
[85] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 53.

backgrounds, as well as women. However, it was not argued that the software used by SWP had that effect, but rather that SWP breached its positive duty to continuously have due regard to eliminate such discrimination. The court stressed that public concern regarding the relationship between the police and BAME communities is significant in today's society. The importance of the Public Sector Equality Duty (PSED) lies in its requirement for public authorities to consider the potential impact of new policies. These policies, while seemingly neutral, may disproportionately affect certain segments of the population.[86]

The lower Divisional Court was convinced by a witness statement that found that there was virtually no difference in the statistics as to gender and race. The Divisional Court, in fact, completely relied on the statement of a police constable without conducting further investigating the claim. Yet, the Court on appeal did not consider the argument sufficient and presented the evidence from an expert witness, Dr Anil Jain, a professor in the Department of Computer Sciences and Engineering at Michigan State University in the United States.[87] Dr Jain explained that the performance of AFR Technology is affected by a number of variables, including 'training datasets'.[88] The witness further claimed that the accuracy of the system depends 'to a considerable extent' on the training dataset and that the systems can suffer from training bias.[89] According to Dr. Jain, SWP was likely unaware of the dataset used to train the AFR system in this situation.Consequently, it would be difficult for SWP to assess whether the AFR system is biassed, as the database statistics, such as the number of men and women, and the different races, would need to be known. The Court then found that the South Wales Police had never attempted to ensure, either directly or through independent verification, that the software program in question is free from unacceptable bias based on race or sex.

Based on the evidence presented, the Court found that South Wales Police could not adequately assess the potential discriminatory impact of the technology. The Court concluded that, given the novelty and controversy surrounding Automatic Facial Recognition, all police forces intending to use this technology in the future must ensure that all reasonable measures are taken to prevent racial or gender bias in the software used.[90]

Netherlands: Syri legislation in breach of European Convention on Human Rights

In February 2020, the District Court of The Hague issued its verdict in the case of NCJM et al. and FNV versus The State of the Netherlands, commonly known as the "SyRI" case. This legal challenge contested the Dutch government's implementation of the Systeem Risico Indicatie, in English System Risk Indication (hereinafter referred to as SyRI), an algorithm aimed at detecting potential instances of social welfare fraud.[91] The court determined that

[86] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 163 to 190.
[87] R Bridges v CC South Wales [2020] EWCA Civ 1058, paras 191-193.
[88] R Bridges v CC South Wales [2020] EWCA Civ 1058, paras 193.
[89] Ibid.
[90] R Bridges v CC South Wales [2020] EWCA Civ 1058, para 194 to para 201.
[91] Rechtbank Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:1878.

both the legislation governing SyRI (Section 65 SUWI Act)[92] and its application failed to meet the standards outlined in Article 8(2) ECHR, which specifies the conditions under which interference with the right to privacy can be deemed necessary and proportionate.

SyRI is a legal instrument used by the Dutch government to detect and combat fraud in the areas of social security and income-dependent programmes, taxes and social security, and labour legislation. SyRI is defined as a technical infrastructure that enables anonymous data analysis and risk reporting in a secure environment.[93] NJCM et al. argue that the SyRI legislation violates individuals' privacy and the right to respect their private lives, as well as lacking necessary safeguards.[94] The claim was brought under Article 8 ECHR, Articles 7 and 8 Charter and Article 17 ICCPR and also a violation of Articles 5, 6, 13, 14, 22 and/or 28 GDPR.[95] The Court only addressed whether the SyRI legislation met the requirements in Article 8 ECHR, while taking into account the general principles of data protection stemming from the Charter and the GDPR.[96]

The Court considered the principles of transparency, the principle of purpose limitation and the principle of data minimisation and concluded that the SyRI legislation was insufficiently transparent and verifiable.[97] Consequently, data subjects would not be able to defend themselves against the fact that a risk report has been submitted about them.[98] The Court argued that the SyRI legislation contained insufficient safeguards to ensure the respect of the right of private life in relation to the risk indicators and the risk model used in SyRI.[99] In light of the lack of insight into the risk indicators and the risk model used in the SyRI system, the Court was unable to assess if the interference with the right to respect for private life was always proportionate and therefore necessary.[100]

It is important to note that no claim was raised regarding the right not to be discriminated against (under Article 14 of the ECHR and/or Article 2(2) of the ICESCR) in connection to the right to privacy, even though the interference specifically concerned citizens of low economic status, which could constitute discrimination on the basis of social origin, property or other status.[101] Yet, the Court recognized that SyRI inadvertently creates links based on bias, such as a lower socio-economic status or an immigration background and, consequently,

---

[92] SyRI is has its legal basis in Section 65 SUWI Act, as amended in January 2024, read in conjunction with Section 64 SUWI Act and Chapter 5a SUWI Deree. Para 4.1.
[93] Syri para 3.1.
[94] Para 6.1.
[95] Para 6.2.
[96] Syri para 6.41.
[97] Para 6.86. See also paras 6.87-6.90.
[98] Para 6.90.
[99] Para 6.95.
[100] Para 6.95.
[101] Adamantia Rachovitsa, Niclas Johann, The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch *SyRI* Case, *Human Rights Law Review*, Volume 22, Issue 2, June 2022, ngac010, https://doi.org/10.1093/hrlr/ngac010.

a risk for discrimination.[102] Moreover, according to the court, to date, SyRI was later only applied in poor neighbourhoods ('problem districts').[103] The court added that 'the right to respect for private life in the context of data processing concerns the right to equal treatment in equal cases, and the right to protection against discrimination, stereotyping, and stigmatisation'.[104] The court was also unable to answer whether the risk of discrimination was sufficiently neutralised.[105]

Koninklijke Maracheusee

On the 14th of February 2023, the Court of Appeals in The Hague delivered a landmark judgement in a case initiated by two citizens, Amnesty International, Controle Alt Delete, RADAR and NJCM-PILP against the Dutch state.[106] The case involved the Dutch Royal Military Police (KMar) using ethnic profiling during their supervision and enforcement of residency status at the border, known as MTV checks (Mobiel Toezicht Veiligheid). During these checks, the KMar can stop individuals crossing the Dutch border to verify their identity, nationality, and residency status. The KMar decides whom to stop based on various 'information-driven'[107] indicators, which sometimes include personal characteristics like skin colour.[108] The KMar argued that these characteristics are always used alongside other factors and are sometimes necessary for identifying a person's nationality or origin during MTV checks.[109]

The applicants in the case argued that using such characteristics amounts to ethnic profiling and discrimination based on race by the KMar and violates the right to non-discrimination under Article 14, and Article 1 of Protocol 12.[110] Furthermore, the applicants demanded that the court impose a ban on the State, including the KMar, from using racial characteristics in the context of MTV checks.[111]

The District Court of The Hague dismissed the applicant's claims, but the Court of Appeal in The Hague ruled in their favour on appeal. The Court of Appeal found that the KMar discriminates based on the protected grounds of race. Given the severe consequences of racial discrimination, such a distinction can only be made for particularly compelling reasons.

---

[102] Syri para 6.9. See also Appelman N, Fathaigh RO and van Hoboken J, 'Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands' (2021) 12 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 257.
[103] Syri para 6.93.
[104] Syri para 6.24.
[105] Ibid. paras 6.49, 6.53, 6.92 6.94.
[106]Gerechtshof Den Haag (2023). ECLI:NL:GHDHA:2023:173. Available online at: https://uitspraken.rechtspraak.nl/#!/details?id=ECLI:NL:GHDHA:2023:173.
[107] *Kmar uitspraak* para 5.2.
[108] *Kmar uitspraak* para 1.1, para 5.1-5.9.
[109] *Kmar uitspraak* para 1.1, para 5.1-5.9.
[110] *Kmar uitspraak* para 6.1, 7.2.
[111] *Kmar uitspraak* para 6.1, 8.9.

The State failed to demonstrate such compelling reasons.[112] This means that KMar was found guilty of racial discrimination, which is prohibited. As a result, the Court of Appeal prohibited the State from making selection decisions in MTV checks that are (partially) based on race.[113]

The Child Care Benefits Scandal

The Dutch tax authority employed a self-learning algorithm to decide on fraudulent child care benefit claims.[114] However, it eventually became clear that the algorithm was systematically classifying double-citizenship claims as high-risk which in turn resulted in officials branding those claims as fraudulent.[115] These mistakes, caused by a discriminatory policy which was translated into the AI system, led to devastating outcomes for affected families with very little explanations as to why.[116] The Netherlands Institute for Human Rights concluded in three separate cases that there was indirect discrimination in the childcare benefit case.[117]

AERIUS case

The AREIUS software was employed in the Netherlands to aid decision-making within the framework of the PAS (Integrated Approach to Nitrogen) program, which is focused on reducing emissions from agricultural activities.[118] In its ruling, the Jurisdictional Division of

---

[112] *Kmar uitspraak* para 8.13-8.15, 8.20-8.21, 8.24.

[113] *Kmar uitspraak* para 9.1-10.2.

[114] See the Dutch Data Protection Authority report on this system at <https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf> accessed 13 May 2024.

[115] Björn ten Seldam, Alex Brenninkmeijer, 'The Dutch benefits scandal: a cautionary tale for algorithmic enforcement' (*eulawenforcement*, 30 April 2021) <https://eulawenforcement.com/?p=7941> accessed 13 May 2024; Fréderike Geerdink, "I cry a lot, every day': victims of the Dutch child benefits scandal fight for compensation' (*Conversationalist.org*, 14 October 2021) <https://conversationalist.org/2021/10/14/i-cry-a-lot-every-day-victims-of-the-dutch-child-benefits-scandal-fight-for-compensation/> accessed 13 May 2024; David Hadwick, 'Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany' (2021) 13(4) World Tax Journal 6; Amnesty International, 'Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal' (*Amnesty.org*, 25 October 2021) <https://www.amnesty.org/en/documents/eur35/4686/2021/en/> accessed 11 November 2023; Netherlands House of Representatives, 'Verslag - Parlementaire Ondervragingscommissie Kinderopvangtoeslag: Ongekend Onrecht' (Report, 17 December 2020) <https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf> accessed 11 November 2023, 7, 14.

[116] Ida Varošanec, 'On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI' (2022) 36(2) International Review of Law, Computers & Technology 95-117.

[117] https://oordelen.mensenrechten.nl/oordeel/2023-101; https://oordelen.mensenrechten.nl/oordeel/2023-102; https://oordelen.mensenrechten.nl/oordeel/2023-103.

[118] ABRvS (Judicial Division of the Council of State) 17 May 2017, ECLI:NL:RVS:2017:1259 ('AERIUS').

the Council of State—the highest administrative court for administrative disputes—argued that AERIUS hindered transparency and access to information for the parties involved, thereby impacting their right to a remedy.[119] Particularly, the Court stated that:

> To avoid this unequal procedural position, in this case the said ministers and the State Secretary are under the obligation to make the choices made, and the data and assumptions used fully, timely and voluntarily public in an appropriate manner so that these choices, data and assumptions are accessible to third parties. This full, timely and adequate disclosure must make it possible to assess, or have assessed, and if necessary contest, with reasons, the choices made and the data and assumptions used, so that real legal protection against decisions based on these choices, data and assumptions is possible, enabling the court to review the lawfulness of these decisions.[120]

Following this decision, the public authority is required to offer transparency regarding the data input, operation, and use of the algorithms that form the basis of the decision. In other words, the obligation to provide justification must be upheld, even when decisions are automated. The Supreme Court applied this reasoning in a separate case involving calculation models used to determine property values for tax purposes.[121]

2.4 National and international AI strategies

2.4.1 National AI strategies

There is a wide variety of national AI strategies aiming to ensure responsible utilisation of AI systems in the public sector. For instance, OECD maintains a live repository of national AI strategies and policies which showcases global efforts to approach the use of AI systems.[122]

[119] See AERIUS judgement, para 14.3.
[120] See AERIUS judgement, para 14.4.
[121] Hooge Raad, 17 August 2018, ECLI:NL:HR:2018:1316, para 2.3.3.
[122] See <https://oecd.ai/en/dashboards/overview> accessed 1 June 2024.

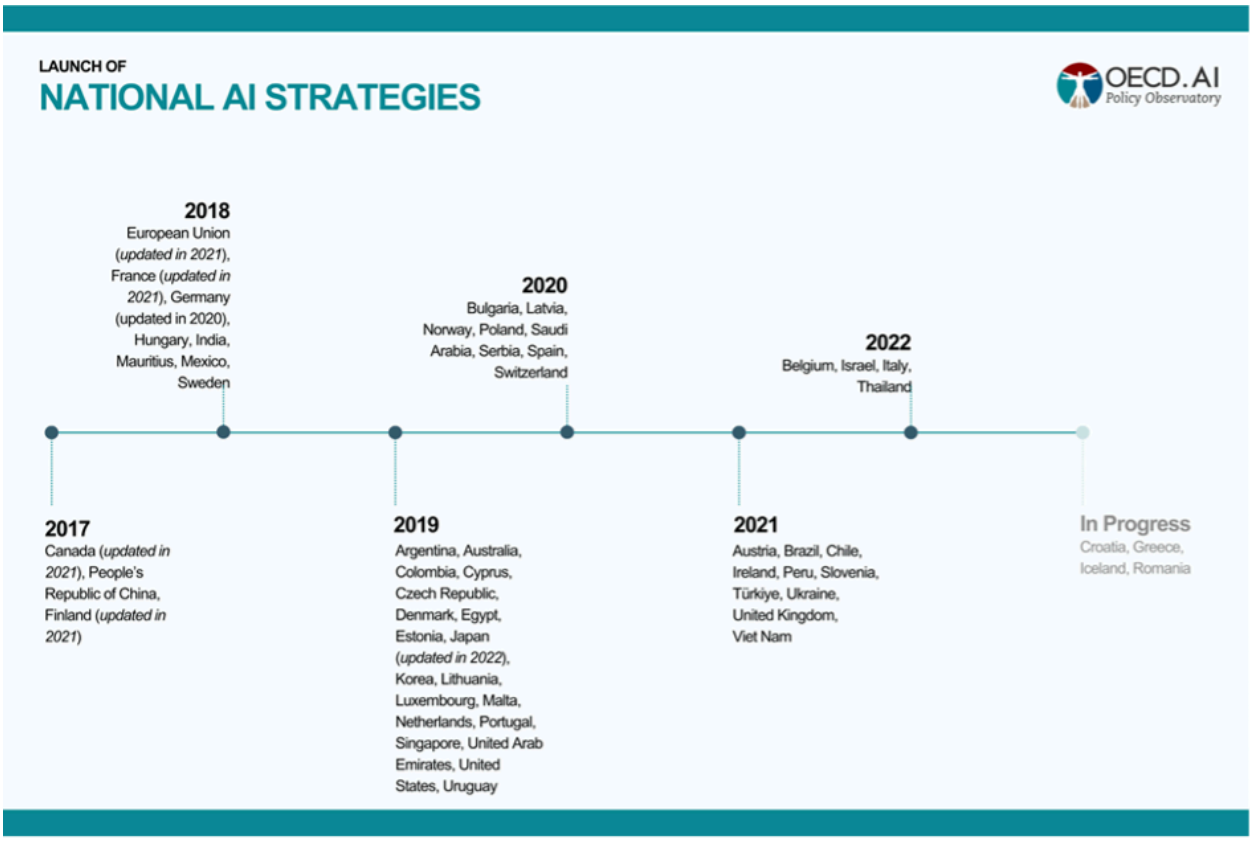


Figure 4. OECD figure depicting the adoption of national AI strategies.[123]

In Europe, several countries have developed or are in the process of developing AI strategies.[124] Majority of them aim at the use of AI in the public sector as a means to improve public services for the benefit of the citizens, coordination of administration and enhance efficiency of routine government processes.[125] For instance, the Netherlands adopted a flexible AI strategy in 2019, involving government collaboration with businesses and academia to address AI challenges.[126] The government aims to use AI for public sector improvement while prioritising ethical considerations and human-centric approaches.

The UK's 2021 AI strategy focuses on three goals: long-term investment in AI, supporting businesses in adopting AI, and ensuring responsible AI governance.[127] The plan includes actions like consultations, research, and ethical frameworks to achieve these goals. Overall,

[123] Lucia Russo, Noah Oder, 'How countries are implementing the OECD Principles for Trustworthy AI' (*OECD.ai*, 31 October 2023) <https://oecd.ai/en/wonk/national-policies-2> accessed 1 June 2024.
[124] R Jorge Ricart, V van Roy, F Rossetti, L Tangi, 'AI Watch – National strategies on Artificial Intelligence: A European perspective' (JRC129123 Report, EUR 31083 EN, 2022); https://ai-watch.ec.europa.eu/index_en>.
[125] Council of Europe, 'Artificial Intelligence in Public Sector' (CAHAI-PDG(2021)03, 11 March 2021) 4.
[126] The Netherlands, 'Strategic Action Plan for Artificial Intelligence' (2019) <https://www.government.nl/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence> accessed 30 November 2023.
[127] HM Government, 'National AI Strategy' (Command Paper 525, September 2021) <https://www.gov.uk/government/publications/national-ai-strategy> accessed 1 December 2023.

the UK aims for a soft-law approach with advisory bodies and existing sectoral regulations, rather than a single overarching AI law.[128]

Belgium adopted a National Convergence Plan for AI development in 2022.[129] This plan aims to make Belgium a leader in AI with nine key objectives. The plan focuses on responsible AI development, promoting business adoption of AI, and using AI for public good in areas like healthcare and sustainability. It also acknowledges the need for education and upskilling the workforce to adapt to the changing job market brought by AI.

France updated its national AI strategy in November 2021, building on the 2018 Villani Report.[130] The strategy focuses on improving data access, investing in AI research and talent, addressing AI's impact on the labour market and environment, and ensuring AI is inclusive and transparent.[131]

Germany's 2018 AI strategy aims to establish Germany as a leader in AI by investing in talent, research, and infrastructure.[132] The strategy prioritises specific sectors like healthcare, environment, and mobility, and focuses on using AI in the public sector. It emphasises open data and developing a regulatory framework for safe and trustworthy AI.[133]

Finland's AI strategy, launched in 2017, is one of the first in Europe.[134] It focuses on both public and private sectors, with ambitious goals like boosting competitiveness and building the world's best public services. The strategy emphasises human-centred AI development and preparing for the changing nature of work.[135] A key aspect of the strategy is the AuroraAI program, which aims to transform public services by providing anticipatory and personalised services to citizens throughout their lives, connecting data from different government

---

[128] Liaison Committee, *AI in the UK: No Room for Complacency* (HL 2019-21, 196) paras 47-61.

[129] AI4Belgium, 'Nationaal convergentieplan voor de ontwikkeling van artificiële intelligentie' (2022) <https://bosa.belgium.be/sites/default/files/content/documents/DTdocs/AI/Nationaal_convergentieplan_voor_de_ontwikkeling_van_artificiele_intelligentie.pdf> accessed 1 June 2024.

[130] Cédric Villani, 'For a Meaningful Artificial Intelligence: Towards a French and European Strategy' (Report 2018) <https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf> accessed 25 November 2023.

[131] See <https://www.aiforhumanity.fr/en/> accessed 25 November 2023.

[132] See <https://www.ki-strategie-deutschland.de/home.html> accessed 29 November 2023.

[133] The Federal Government, 'Artificial Intelligence Strategy of the German Federal Government: 2020 Update' (Report, December 2020) <https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf> accessed 29 November 2023.

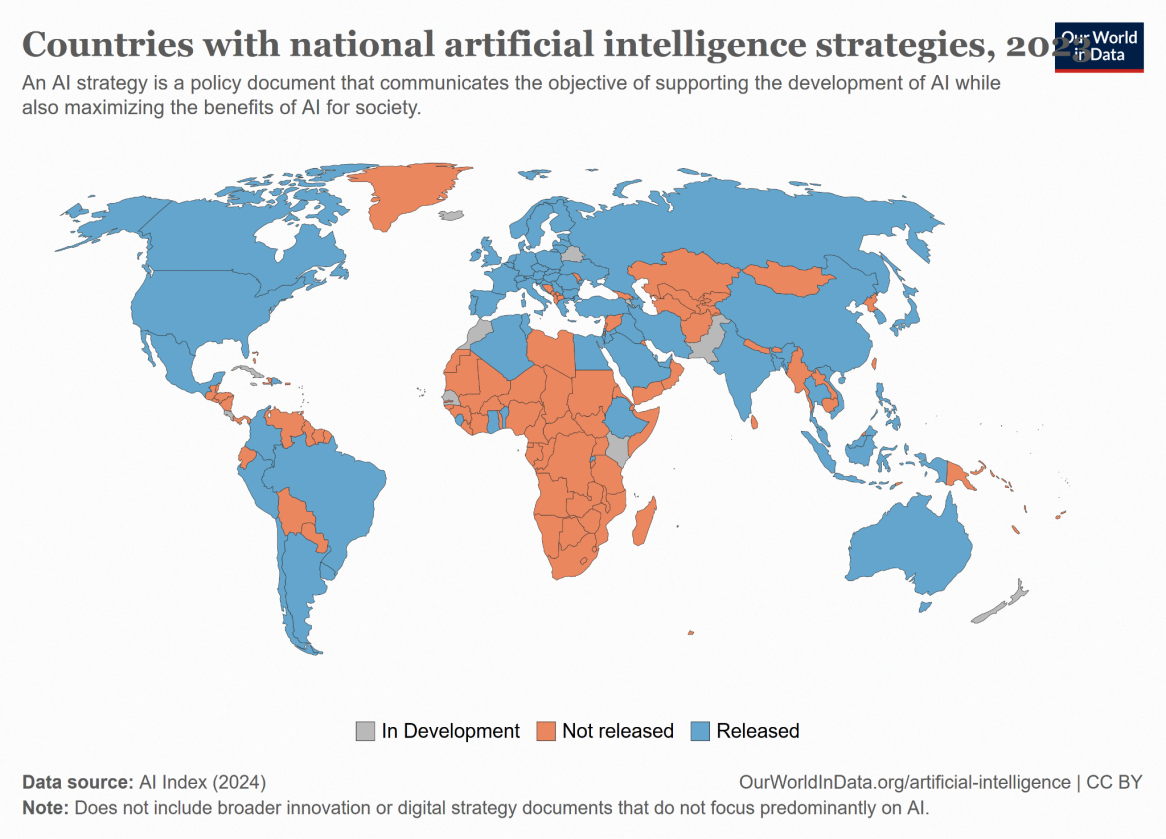
[134] Gianluca Misuraca, C van Noordt, 'Overview of the use and impact of AI in public services in the EU' (JRC120399, EUR 30255 EN, 2020) 60.

[135] Ministry of Economic Affairs and Employment, 'Finland's Age of Artificial Intelligence' (Publications of Ministry of Economic Affairs and Employment 47/2017, Helsinki 2017) <https://julkaisut.valtioneuvosto.fi/handle/10024/160391> accessed 24 November 2023; Ministry of Economic Affairs and Employment, 'Leading the way into the age of artificial intelligence' (Publications of Ministry of Economic Affairs and Employment 2019:41, Helsinki 2019) <https://julkaisut.valtioneuvosto.fi/handle/10024/161688> accessed 24 November 2023.

agencies to create a network of AI applications, and empowering citizens to manage their data (DigiMe) and access personalised services.[136] These are merely some of the examples of national AI strategies, illustrating that AI uptake in various sectors inspired initiatives for ensuring trustworthy AI on national levels.

2.4.2 International level

Many strategies have been developed in other parts of the world as well. For instance, Canada established its national AI strategy in 2017[137] and the US has also become quite active with the introduction and legislative proposals for various aspects of AI systems.[138]



[136] See Ministry of Finance, 'AuroraAI – Towards a human-centric society' (March 2019) <https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019–2023.pdf> accessed 24 November 2023.

[137] https://cifar.ca/ai/.

[138] Nestor Maslej, Loredana Fattorini, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, and Jack lark, "The AI Index 2024 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024, 391; See also the Executive Order issued in 2023 at: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Figure 5. Countries with national AI strategies.[139]

International organisations such as the World Economic Forum, UN, Amnesty International and OECD have also issued their own guidelines for matters concerning algorithmic design.[140]

2.5 Ethical instruments and Soft-law/policy framework

Ethical instruments play an important role in ensuring trustworthy AI uses. They are an important tool for building upon the existing regulatory framework.[141] Although the list is non-exhaustive, we name a few here.

The Algorithm Framework (Algoritmekader)

In response to, and in order to prevent, cases like the Child Care Benefits Scandal and SyRI, the Dutch Ministry of Internal Affairs is developing a framework to support organisations in the responsible development and deployment of algorithmic systems.[142] The *Algoritmekader* offers practical normative guidance to organisations on how to meet and operationalize minimum regulatory requirements, and govern risks to fundamental rights throughout the lifecycle of algorithmic systems.[143] The minimum regulatory requirements stem from current legal instruments such as European human rights legislation, as well as emerging legal instruments such as the Artificial Intelligence Act.[144] This practical guidance aims to bring together, and is based on, best practices, use cases, and feedback from end users and supervisory authorities.[145] The *algoritmekader* enables governmental organisations to see which

---

[139] AI Index (2024) – with minor processing by Our World in Data. "Countries with national artificial intelligence strategies" [dataset]. AI Index, "AI Index Report" [original data]. Retrieved August 30, 2024 from https://ourworldindata.org/grapher/national-strategies-on-artificial-intelligence.

[140] World Economic Forum, 'Guidelines for AI Procurement' (2019) https://www.weforum.org/whitepapers/ai-government-procurement-guidelines ; General Comments from human rights treaty monitoring bodies, e.g. GC No. 25 of the Committee on Economic, Social and Cultural Rights; GC No. 25 of the Committee on the Rights of the Child; United Nations Educational, Scientific and Cultural Organization, 'UNESCO Recommendation on the Ethics of Artificial Intelligence' (2021) https://unesdoc.unesco.org/ark:/48223/pf0000381137; United Nations Educational, Scientific and Cultural Organization, 'UNESCO Recommendation on the Ethics of Artificial Intelligence' (2021) https://unesdoc.unesco.org/ark:/48223/pf0000381137; Amnesty International and Access Now, 'Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems' (2018) https://www.torontodeclaration.org/; Information Technology Industry Council, 'Global AI Policy Recommendations' (2021) https://www.itic.org/documents/artificial-intelligence/ITI_GlobalAIPrinciples_032321_v3.pdf; Asilomar AI Principles (11 August 2017) https://futureoflife.org/open-letter/ai-principles/.

[141] European Commission, 'Building Trust in Human-Centric Artificial Intelligence' COM (2019) 168 final.

[142] https://minbzk.github.io/Algoritmekader/.

[143] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/algoritmes/algoritmerekader/; https://minbzk.github.io/Algoritmekader/.

[144] https://minbzk.github.io/Algoritmekader/vereisten/.

[145] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/algoritmes/algoritmerekader/; https://minbzk.github.io/Algoritmekader/.

requirements and risk governance measures, roles, best practices, and instruments are relevant to each stage of the life cycle.[146]

Fundamental Rights Algorithm Impact Assessment

The Fundamental Rights and Algorithm Impact Assessment (FRAIA) functions as a discussion and decision-making aid within governmental organisations developing or using algorithms to facilitate ex-ante compliance with fundamental rights and was published in 2021 by the Dutch Ministry of the Interior and Kingdom Relations. This tool, composed of questions to be discussed, facilitates interdisciplinary dialogue between relevant stakeholders seeking to identify and mitigate risks to fundamental rights.[147] Concerning the design of algorithmic systems, the FRAIA contains questions about the rationale for the use of an algorithm as well as its input. In particular, questions are asked about the underlying reasoning for its development and use, public values at stake, intended effects, and objectives of the algorithmic system. Furthermore, questions are asked about the 'shape' of the algorithmic system. This includes questions about the algorithm, the input data, the embedded preconditions, and transparency and explainability.[148]

Z inspection

The Z-Inspection self-assessment process uses the EU guidelines for trustworthy AI and applied ethics to evaluate the trustworthiness of AI systems.[149] In this regard, the process focuses on the identification and discussion of ethical issues and tensions throughout socio technical situations.[150] In the design stage, the Z-Inspection process can be used to generate insight into how to design and develop trustworthy systems.[151] Sociotechnical scenarios relating to the use and effects of a system are built and discussed by an interdisciplinary group to describe the system and highlight ethical issues and tensions. Based on this process, tensions, and issues are resolved and recommendations are provided to stakeholders.[152]

Handbook Non-Discrimination by Design

The non-discrimination by design handbook assists teams in building AI systems that comply with the right to non-discrimination. This tool helps teams building AI systems to identify

---

[146] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/algoritmes/algoritmerekader/; https://minbzk.github.io/Algoritmekader/.

[147] https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms.

[148] https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms Pages 6-46.

[149] https://z-inspection.org/the-process-2/.

[150] https://z-inspection.org/how-to-assess-trustworthy-ai-in-practice/.

[151] https://arxiv.org/ftp/arxiv/papers/2206/2206.09887.pdf.

[152] https://arxiv.org/ftp/arxiv/papers/2206/2206.09887.pdf.

different stages in the design process, bring together stakeholders, and discuss relevant questions from a technical, legal, and organisational perspective.[153] Additionally, the handbook uses several scenarios to illustrate challenges and demonstrate how those challenges can be addressed.[154] The handbook is structured along various stages within the design of a system, highlighting questions, challenges, perspectives, and scenarios within each stage.[155]

Algoprudence

Algoprudence refers to case-based normative advice for ethical algorithms based on inclusive, diverse, and deliberative commissions set up by Algorithm Audit. While Algorithm Audit cannot issue legally binding rulings, its non-binding ethical advice can serve as guidance for the responsible design and deployment of algorithms.[156]

OECD Principles for Trustworthy AI

Organisation for Economic Co-operation and Development (OECD) has developed AI Principles for trustworthy AI which are accompanied by the Tools for trustworthy AI report.[157] These value-based principles include inclusive growth, sustainable development and well-being; respect for human rights and democratic values, including fairness and privacy; transparency and explainability; robustness, security and safety; and accountability. They aim to guide AI actors and policymakers in enabling trustworthy AI. When it comes to the implementation, the toolkit consists of technical approaches consisting of software tools, technical documentation, certification and standards, product development or lifecycle tools, and technical validation tools.[158]

Council of Europe

---

153 https://www.tilburguniversity.edu/sites/default/files/download/04%20handbook%20non-discrimination%20by%20design%28ENG%29.pdf.

154 https://www.tilburguniversity.edu/sites/default/files/download/04%20handbook%20non-discrimination%20by%20design%28ENG%29.pdf.

155 https://www.tilburguniversity.edu/sites/default/files/download/04%20handbook%20non-discrimination%20by%20design%28ENG%29.pdf.

156 https://algorithmaudit.eu/algoprudence/how-we-work/; Meuwese A, Parie J and Voogt A, 'Hoe "Algoprudentie" Kan Bijdragen Aan Een Verantwoorde Inzet van Machine Learning-Algoritmes' (2024) NJB 2024/556 Nederlandse Juristenblad .

157 OECD, 'Recommendation of the Council on Artificial Intelligence' (adopted 22 May 2019, amended on 8 November 2023) OECD/LEGAL/0449; OECD, 'Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems' (Report, OECD Digital Economy Papers No 312, 2021) <https://doi.org/10.1787/008232ec-en> accessed 23 April 2024.

158 OECD, 'Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems' (Report, OECD Digital Economy Papers No 312, 2021) <https://doi.org/10.1787/008232ec-en> accessed 23 April 2024, 10.

The Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law, issued on 7 July 2023, proposes a comprehensive regulatory framework for AI systems.[159] Chapter III of the draft convention outlines key principles for the design, development, use, and decommissioning of AI systems. Article 7 stresses the importance of transparency and oversight, underscoring the need for AI processes to be understandable and subject to scrutiny, which developers should consider during the design phase. Article 9 focuses on equality and non-discrimination, mandating that AI systems be free from biases, also to be addressed during the design stage. Additionally, Article 10 highlights privacy and personal data protection, ensuring alignment with existing data protection frameworks to safeguard individuals' privacy rights in AI applications.

European Parliament

European Parliament Resolutions such as those from October 2020, provide additional guidance and recommendations on ethical aspects of AI and civil liability regimes.[160] They emphasise the importance of upholding fundamental rights, including privacy, non-discrimination, and accountability, in the development and deployment of AI technologies.[161] These resolutions contribute to the ongoing discourse on AI governance and help shape policies aimed at promoting responsible AI innovation while safeguarding individual rights and societal values. Point 5 shares the importance of incorporating Union principles into the design of new technologies, yet it does not go into details on whether this would entail.

2.6 AI Liability In The Public Sector

In the current EU legal landscape, AI liability in the public sector is hard to define in specific terms and legislation. In the absence of proper EU legislation, there is no clear way to determine the form of liability or identify the person responsible for the damage. Clarity on this issue may only be gained by bringing a preliminary reference to the CJEU and starting the arduous process of case precedent.

---

[159] Committee on Artificial Intelligence (CAI), Consolidated Working Draft of the Framework Convention on Artificial Artelligence, 7th July 2023, CAI(2023)18, https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66.

[160]European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL) https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html ;
European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020IP0276.

[161] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)) https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html.

Considering the particularities of each member state's national law and the evolving case law on state liability and accountability, this chapter will focus on analysing several methods of applying design liability to AI systems employed in the public domain. Firstly, it will discuss the current state of AI design liability in contract and tort law, as well as the liability gaps that arise. Secondly, it will explain different methods of applying AI liability, whether defining the AI system as a legal actor or as a tool, in order to bridge liability gaps. Lastly, it will also analyse the proposed EU AI Liability Directive, as well as the updated Product Liability Directive, and assess their regulation of liability and protection of individuals.

2.6.1 Civil AI Liability

In the absence of adequate liability legislation, liability in the public sector is assessed purely on the coattails of other legal domains. For instance, the Dutch Government's child care benefit scandal[162] was sanctioned on the basis of data protection law,[163] and in the similar SyRI case the use of the unfair AI system was banned based on the ECHR.[164] Both of these cases are prime examples of faulty AI design and the urgent need for liability guidelines for public institutions to address future cases.

By comparison, liability in civil law is a bit easier to determine:[165] either contractual liability or tort liability may offer an effective remedy in certain cases. Contractual liability would require specific provisions in the contract mentioning and governing the use of AI systems.[166] For tort liability, most Member States 'contain a regulatory mix' between fault-based and strict liability.[167] While fault-based liability is the default type, it does not apply in cases of poor AI design.[168] When talking about algorithmic design especially, the blame does not fall on the

---

[162] Heikkilä, Melissa. "Dutch scandal serves as a warning for Europe over risks of using algorithms." Politico, March 29, 2022. https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms.

[163] Agrawal, Ananaya. "Netherlands data protection watchdog finds tax authority violated GDPR." Jurist, October 31, 2021. https://www.jurist.org/news/2021/10/netherlands-data-protection-watchdog-finds-tax-authority-violated-gdpr.

[164] van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgement. *European Journal of Social Security*, *23*(4), 323-340. https://doi.org/10.1177/13882627211031257.

[165] De Conca, S. (2022). Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-empowering Solutions. In: Custers, B., Fosch-Villaronga, E. (eds) Law and Artificial Intelligence. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-523-2_13.

[166] Đurić, L., & Ignjatović, J. "Who's Responsible? Addressing Liability in the Age of AI." Gecic Law. June 22, 2023. https://geciclaw.com/whos-responsible-addressing-liability-in-the-age-of-artificial-intelligence/.

[167] Martin Ebers, Liability For Artificial Intelligence And EU Consumer Law, 12 (2021) JIPITEC 204 para 1. https://www.jipitec.eu/issues/jipitec-12-2-2021/5289.

[168] da Fonseca, A.T., Vaz de Sequeira, E., Barreto Xavier, L. (2024). Liability for AI Driven Systems. In: Sousa Antunes, H., Freitas, P.M., Oliveira, A.L., Martins Pereira, C., Vaz de Sequeira, E., Barreto Xavier, L. (eds)

human element, but rather the product. Therefore, it would be not desirable to require a proof that the person in charge did not fulfil their duty of care and thus fulfil the level of negligence[169] required by European tort systems.[170]

This is why strict liability proves exceptionally useful, as it often represents the only effective remedy in cases of design issues. Product liability does not require any negligence or intent, but rather that the product itself is flawed. While this seems like an easy fix for AI liability, in reality it is much more difficult, especially since in strict liability, the consumer bears the burden of proof of the product's defect.[171] Because of AI's black-box nature, it can be hard or even impossible to label the product as defective,[172] even with considerable expertise. This is partly because an AI system contains numerous different software and hardware parts that all interact with each other. Not only is it difficult to prove which part of the AI is flawed,[173] but sometimes the problem may be caused by the way the parts interact with one another.[174] Another issue caused by the black-box effect is that it makes it hard to ascertain a specific blame, especially in cases with multiple actors working on different parts of the AI,[175] such as the software, or the hardware, or the training dataset. As AI changes over time and is not a static product, it is difficult to ascertain the faulty actors and their degree of blame, and furthermore to establish a causal link.[176]

Multidisciplinary Perspectives on Artificial Intelligence and the Law. Law, Governance and Technology Series, vol 58. Springer, Cham. https://doi.org/10.1007/978-3-031-41264-6_16.

[169] Zech, H. Liability for AI: public policy considerations. *ERA Forum* 22, 147–158 (2021). https://doi.org/10.1007/s12027-020-00648-0.

[170] European Commission, Directorate-General for Justice and Consumers, Karner, E., Koch, B., Geistfeld, M., *Comparative law study on civil liability for artificial intelligence*, Publications Office of the European Union, 2021, https://data.europa.eu/doi/10.2838/77360 p. 39.

[171] De Conca, S. (2022). Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-empowering Solutions. In: Custers, B., Fosch-Villaronga, E. (eds) Law and Artificial Intelligence. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-523-2_13.

[172] Padovan, P.H., Martins, C.M. & Reed, C. Black is the new orange: how to determine AI liability. Artif Intell Law 31, 133–167 (2023). https://doi.org/10.1007/s10506-022-09308-9.

[173] Vasudevan, Amrita. "Addressing the Liability Gap in AI Accidents," Centre for International Governance Innovation Policy Brief No. 177, July 31, 2023. https://www.cigionline.org/publications/addressing-the-liability-gap-in-ai-accidents/

[174] De Conca, S. (2022). Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-empowering Solutions. In: Custers, B., Fosch-Villaronga, E. (eds) Law and Artificial Intelligence. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-523-2_13.

[175] Vasudevan, Amrita. "Addressing the Liability Gap in AI Accidents," Centre for International Governance Innovation Policy Brief No. 177. July 31, 2023. https://www.cigionline.org/publications/addressing-the-liability-gap-in-ai-accidents/.

[176] De Conca, S. (2022). Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-empowering Solutions. In: Custers, B., Fosch-Villaronga, E. (eds) Law and Artificial Intelligence. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-523-2_13.

There are techniques being developed for a better understanding of AI which would help minimise those issues, such as XAI (Explainable Artificial Intelligence).[177] However, those techniques are far from being finalised and usable in court settings at present. It is therefore clear that in its current state, civil liability leaves a lot of loopholes, and requires further legislation and reglementation.

2.6.2 AI Liability Directive and Product Liability Directive

Another issue with product liability in AI design is that it is still unclear whether AI is a product and included in product liability laws, or a service. The revised EU Product Liability Directive (PLD) clarifies this issue by broadening the scope of the Directive to explicitly include software, thus allowing product liability to apply. Furthermore, the revised Directive also allows holding software and digital service providers liable, straying from the customary viewpoint that only manufacturers and importers can be held liable.[178] This shift is a positive development, considering that an AI system is a complex technological ecosystem built by different actors at different stages of its lifetime.[179] As a result, it is often difficult, if not impossible, to attribute the fault of a defective product to a single manufacturer.

The PLD serves as a complementary piece of legislation to the AI Liability Directive (AILD). While the PLD governs strict liability, the AILD focuses on fault-based liability. In its revised form, the PLD introduces significant improvements to AI liability law. Notably, it allows for holding providers liable based on a product defect. Key advancements include the introduction of a presumption of defectiveness (Article 9(2)) and a presumption of a causal link between the defect and the damage (Article 9(3)).[180] The presumption of defect can be triggered by the provider's failure to disclose system information, or the claimant showing either an obvious malfunction, or a violation of safety rules on the controller's part. The presumption of the causal link can be triggered if the damage is 'typically consistent' with the defect.[181]

The existence of those presumptions in the Directive certainly simplifies and clarifies the process of proving a flaw in the AI system. As previously mentioned, with current legislation

---

[177] Padovan, P.H., Martins, C.M. & Reed, C. Black is the new orange: how to determine AI liability. Artif Intell Law 31, 133–167 (2023). https://doi.org/10.1007/s10506-022-09308-9.
[178] Cameron, S., Seinen, W. "AI liability addressed in fresh EU proposals," Pinsent Masons, October 3, 2022. https://www.pinsentmasons.com/out-law/news/ai-liability-addressed-eu-proposals.
[179] Ibid.
[180] Samar Abbas Nawaz, "The Proposed EU AI Liability Rules: Ease or Burden?" European Law Blog 48/2022, November 7, 2022 https://europeanlawblog.eu/2022/11/07/the-proposed-eu-ai-liability-rules-ease-or-burden/.
[181] Ibid.

proving an AI system is defective proves to be incredibly hard, partly from the lack of understanding of the system. While the new version of the Directive helps on this front, it is necessary to observe that the burden of proof still falls largely on the consumer, and not on the system provider. Furthermore, Recital 34 also gives the producer the possibility to show that the system does not contain any 'excessive technical difficulties', thus rebutting the claim. Considering the black box effect, it is unclear how much these provisions actually help consumers in practice: the Directive does not do much to help lift such a heavy burden of proof. Most importantly, the PLD only allows compensation in cases of 'material loss' (art 4(6)).[182] Therefore, users harmed through faulty or discriminatory AI systems such as credit-scoring algorithms or the previously banned fraud-detecting algorithm SyRi[183] would be unable to claim compensation under the PLD, and instead need to prove fault through the AILD.

This Directive, by contrast, is the EU's main designated piece of legislation on AI liability. Regulating non-contractual civil liability rules, the AILD is focused on fault-based liability, while the revised PLD fills in the gaps for strict liability[184]. Just like the PLD, the AILD also aims to help consumers through presumptions, namely the presumption of fault, and the presumption of causality (article 3, article 4).[185]

In order to enjoy the presumption of causality, a claimant must firstly prove three key conditions. Firstly, the provider's or user's fault must be either proven or presumed by a court. Secondly, it should be reasonably likely, considering the case's circumstances, that the fault influenced the output or failure of the AI system to produce an output. Lastly, claimants must demonstrate that the AI system's output or failure led to the incurred damage.

Therefore, the AILD not only holds liable the manufacturer of the system, but also its current user: however, they have different fault requirements. In cases of high-risk AI systems[186], a user, such as a public body, would only be at fault if they violated the usage

---

[182] Ibid.

[183] Hussain, Hafsah. "Dutch Court Finds SyRI Algorithm Violates Human Rights Norms In Landmark Case." Human Rights Pulse, March 22, 2020. https://www.humanrightspulse.com/mastercontentblog/dutch-court-finds-syri-algorithm-violates-human-rights-norms-in-landmark-case..

[184] Samar Abbas Nawaz, "The Proposed EU AI Liability Rules: Ease or Burden?" European Law Blog 48/2022, November 7, 2022 https://europeanlawblog.eu/2022/11/07/the-proposed-eu-ai-liability-rules-ease-or-burden/.

[185] Wagner, Gerhard. "Liability Rules for the Digital Age: – Aiming for the Brussels Effect –" *Journal of European Tort Law*, vol. 13, no. 3, 2022, pp. 191-243. .

[186] Ibid.

instructions or exposed the system to irrelevant input data (article 4(3))[187]. Those accusations however still need to be proven by the claimant.

The AILD does aim to ease the proof process for fault-based liability. Article 3, for example, allows an injured person to seek information about the AI system through the national court, and further incentivises disclosure of relevant information by providing for the presumption of causality to be rebutted[188] in circumstances where a user can demonstrate that "sufficient evidence and expertise is reasonably accessible for the claimant to prove the causal link".

Despite those provisions, issues still remain, as it is consistently difficult for claimants to prove a presumption of causality. In this case, the claimant faces a heavy burden of proof: they must provide evidence in support of the plausibility of the claim, must identify non-compliance, and must demonstrate a link between the AI system's action and the damage sustained, article 4(3(1)). Additionally, the AILD is mainly aimed at manufacturers[189] and does little for user liability, setting a limited duty of care, even for high-risk systems.

2.6.3 Liability Gaps. AI as Actor vs AI as Tool[190]

While comparatively an improvement from currently applicable tort law rules, the PLD and AILD offer little procedural ease to injured parties, and do not impose sufficient obligations on users, leaving a significant gap in regulatory measures for user liability. Considering those remaining liability gaps, there are multiple liability guidelines that could be used to bridge them.

One of those liability suggestions is based on the AI system's capacity to make its own autonomous decisions. The European Parliament proposed in a 2017 resolution that AI be treated as a legal entity,[191] an 'electronic person' (e-person) if only for the purposes of liability. However, this liability would only make sense if the AI was able to economically satisfy damages claims.[192] Therefore, the AI systems would be allocated a fund specifically used to fulfil liability payments. This may not be the best solution for a public institution, as this

---

[187] Ibid.

[188] Cameron, S., Seinen, W. "AI liability addressed in fresh EU proposals," October 3, 2022. Pinsent Masons. https://www.pinsentmasons.com/out-law/news/ai-liability-addressed-eu-proposals. .

[189] Wagner, Gerhard. "Liability Rules for the Digital Age: – Aiming for the Brussels Effect –" *Journal of European Tort Law*, vol. 13, no. 3, 2022, pp. 191-243..

[190] Paulius Čerka, Jurgita Grigienė, Gintarė Sirbikytė, Liability for damages caused by artificial intelligence, Computer Law & Security Review, Volume 31, Issue 3, 2015, Pages 376-389..

[191] Ibid.

[192] Beckers, A., & Teubner, G. (2022). *Three liability regimes for artificial intelligence: Algorithmic actants, hybrids, crowds.* Hart Publishing.

solution implies a heavy financial burden and allows for a partial externalisation of the harm, as any damage exceeding the liability fund would simply remain with the victim.[193]

The opposite has also been suggested: that AI remains a tool and users become fully liable for its mistakes. This approach would function within the framework of vicarious liability, based on the Roman law of *respondeat superior* - let the master answer.[194] Traditionally used in the context of employment in both civil and common law systems, this doctrine would see the AI as acting on behalf of a user, and therefore render the user of the AI system liable for any of its failures. Users could, of course, claim the AI manufacturers for damages in cases of faulty or defective AI.[195] As to public institutions, it would be much easier for them to fulfil the causality presumptions required by legislative acts such as the AILD and the PLD.
Vicarious liability would not only prove useful in situations of fault-based liability, but would also bridge the previously discussed liability gap,[196] when there is a defect, but it is either impossible to trace back to a legal subject, or there are too many manufacturers involved to share liability.

2.6.4 Closing Gaps. Article 47 CFR

One legal solution in closing liability gaps can be found in the Charter of Fundamental Rights of the European Union (CFR). As Article 47 of the CFR, Right to an effective remedy and to a fair trial, has been granted direct effect ever since the *Egenberger* case,[197] this article may be used in closing certain liability gaps. By lending its direct effect to certain articles of the AILD and PLD, Article 47 would help establish liability in situations where either the law or the technology is not sufficiently clear. Specific articles that could benefit from direct effect are the previously discussed Article 3 and 4 of the AILD. Article 3, Disclosure of evidence and rebuttable presumption of non-compliance, would have strengthened value in cases where the Directive is not properly implemented or national legislation toughens the process of disclosure. This article is closely connected to Article 47 CFR, given its link with the right to fair trial and equality of arms. Similarly, Article 4, dealing with the rebuttable presumption of a causal link in the case of fault, would be improved through the application of Article 47 CFR. One of the main liability issues caused by AI is their high complexity, making it difficult for claimants to properly prove the causal link, even with the right

[193] Wagner, Gerhard. "Liability Rules for the Digital Age: – Aiming for the Brussels Effect –" *Journal of European Tort Law*, vol. 13, no. 3, 2022, pp. 191-243. .
[194] Čerka, P., Jurgita G., and Gintarė S., "Liability for damages caused by artificial intelligence." *Computer Law & Security Review*, vol. 31, no. 3, June 2015, pp. 376-389 .
[195] Ibid.
[196] Glavaničová, D., Pascucci, M. Vicarious liability: a solution to a problem of AI responsibility?. *Ethics Inf Technol* 24, 28 (2022), .
[197] Case C-414/16 ECLI:EU:C:2018:257. .

information available. This could compromise the right to effective remedy provided by Article 47 CFR, thus proving its importance.

2.6.5 Conclusion

It is hard to discuss AI Liability in a legal context that is undetermined. Current liability law leaves space for many liability gaps, from lack of well-defined legislation to challenges in comprehending AI processes and identifying flaws. Unfortunately, the AI Liability and Product Liability Directives do not manage to fully fill those gaps, and make it hard on claimants to fulfil presumptions of causality by introducing a heavy burden of proof. Furthermore, the Directives have quite a limited range of scope of applications for public institutions: as users, they can only truly be held liable when violating usage instructions or outright tampering with the system, leaving aside occasions where despite proper use, the AI still ends up harming numerous individuals. Taking liability gaps into account, the best solution for liability of the public administration seems to be the approach of vicarious liability. By assigning the user liable in cases of malfunction or improper use, public bodies are bound to take more care when implementing and employing AI systems.

2.7. Concluding remarks: Algorithmic design stage

In the design stage of the AI lifecycle, public entities must prioritise a deep understanding of the problem they aim to solve before committing to AI as a solution. The ideation phase of design thinking is crucial here, as it encourages a thorough evaluation of whether AI is truly the best tool for addressing the identified issue (*technology impact assessment*). Rushing into AI implementation without this understanding can lead to misaligned solutions that do not effectively address the problem. Therefore, taking the time to properly analyse the situation is essential for ensuring that the decision to use AI is both appropriate and beneficial.

Once the decision to use AI has been made, the design stage involves several critical choices that can significantly impact the outcomes of the system. These decisions include determining which data to collect, selecting the appropriate algorithm, and defining the variables and indicators that the system will use. Each of these choices carries potential risks, particularly if they are not carefully considered. For example, using biassed training data or discriminatory predictive indicators can result in outcomes that violate fundamental rights, such as equal treatment and non-discrimination as seen in the Syri case and the Koninklijke Maracheusee cases. Therefore, it is imperative to identify and mitigate any design flaws early in the process to avoid unintended negative consequences.

Furthermore, compliance with legal frameworks like the GDPR, AI Act, and relevant case law is essential throughout the design stage to protect individuals' rights (*Data Protection Impact Assessment, Privacy Impact Assessment*). These regulations provide guidance on critical aspects

such as data protection, transparency, and accountability, which are crucial for preventing harmful outcomes. By following these legal standards and rigorously assessing the potential risks associated with AI design, public entities can develop AI systems that are not only effective but also aligned with ethical principles and fundamental rights. This careful and deliberate approach to AI design ensures that the technology serves the public interest while safeguarding against discrimination and other adverse effects.

2.7.1 Guiding questions

The aim of this handbook is to equip participants with practical information and real-world case studies to help them mitigate potential conflicts with fundamental rights during AI system development. In this section, we highlight some of the most critical questions that public entities should consider during the algorithm design stage, ensuring that their AI initiatives are both effective and aligned with legal and ethical standards.

Fundamental Rights

- Which fundamental rights, such as non-discrimination, equal treatment, privacy, data protection, and human dignity, are potentially affected by this system?
- Could the design of the algorithm unintentionally embed biases?
- Are there discriminatory risk indicators or criteria that could lead to direct or indirect discrimination, especially for vulnerable groups like women with children?
- Does the algorithm's design respect human dignity, or does it reduce individuals to mere data points or scores?
- How does the system ensure fair treatment of all individuals, regardless of their personal circumstances?

Transparency

- How transparent/accessible/public are the criteria, weighting, and cut-offs used in the algorithm?
- Is the system design clear and understandable to ensure that individuals have access to effective remedies if their rights are infringed?

Impact Assessments:

- Does the system involve the collection and processing of sensitive data categories?
- Have comprehensive Data Protection Impact Assessments (DPIA) and Privacy Impact Assessments (PIA) been conducted to identify and mitigate privacy risks?
- Has a thorough Fundamental Rights and Technology Impact Assessment (FRIA) been conducted to evaluate potential adverse impacts on citizens and their fundamental rights?

3. Organisational implementation/embedding

Lorena Flórez, Lucas Haitsma, Yusrah Aminu Ibrahim, Linda Ardenghi, Greta Berardi, Alma Onaie

3.1 Risks for fundamental rights in the implementation of AI systems used in public administration

As organisations increasingly rely on algorithmic decision-making systems in public administration, it is crucial to understand the key issues that can arise throughout the lifecycle of these systems — *from the initial design to their organisational implementation and the final decision-making process*—. Flaws in the algorithmic design, such as biassed training data or opaque 'black box' models, can lead to outcomes that clash with fundamental rights. Similarly, how these systems are embedded within an organisation's workflows and the quality of human oversight can significantly impact the fairness and accountability of the decisions.

To address these challenges, a range of legal, regulatory, and ethical frameworks have emerged to guide the responsible development and deployment of AI systems. Principles of transparency, explainability, and rigorous impact assessments, as outlined in initiatives like the EU's Artificial Intelligence Act and the OECD's Principles on AI, provide important guardrails for organisations implementing algorithmic decision-making tools in the public sector. Understanding and applying these frameworks can help mitigate the risks associated with the organisational implementation and embedding of AI systems, ensuring they serve the public good while respecting individual rights and liberties.

This section of the handbook will provide a comprehensive examination of the implementation stage in the AI life cycle, with a particular focus on its implications for fundamental rights. We will begin by explaining the implementation stage, detailing its components and the technological issues involved. This will be followed by an in-depth exploration of fundamental rights issues that arise during implementation, illustrated through real-world examples and case studies from various jurisdictions. Additionally, it will state legal, regulatory, and ethical frameworks governing the implementation of AI systems in public administration. This analysis will be structured across multiple levels. Particular attention will be given to ethical principles, such as transparency, and how these are reflected in soft law and policy frameworks across different governance levels. Furthermore, this section will address the critical issues of liability and effective remedy in the context of AI implementation, providing insights into how legal systems are adapting to these new challenges. Finally, we will offer practical guidelines for judges, aimed at assisting them in safeguarding fundamental rights during the implementation of AI systems in public administration.

3.2 Implementation stage of AI life cycle

The implementation stage of the AI life cycle is a critical phase that involves the deployment and integration of the developed AI system within an organisation's operations and

processes. This stage encompasses several key technological, organisational and governance considerations and challenges.
Firstly, from a technical perspective, the integration of the AI system into the existing technological infrastructure *(interoperability)* of the organisation is a crucial step.[198] This includes ensuring seamless data exchange, secure communication, and the compatibility of the AI system with other software and systems used within the organisation. Proper system integration is essential to ensure the smooth and efficient functioning of the AI-powered tools *(scalability)*.[199]

Furthermore, from an organisational point of view, the implementation stage also involves the configuration and customization of the AI system to align with the specific needs and requirements of the organisation *(alignment)*. This may include adjusting the AI models, fine-tuning the algorithms, and optimising the system's performance for the intended use cases. Attention must be paid to understand the different functionalities that the system could have. In the same line, it is important for the organisation to identify their limitations on applying these tools. For instance, the most common functionalities can be divided as follows:

Functionality	Description	Example
Classification	It is a supervised learning technique used to predict a discrete, categorical output value based on one or more input variables. It is used to assign data points to specific predefined classes or categories. Thus, based on probabilities, AI systems can classify information.	An email server that automatically filters all incoming emails and tells us which ones are not so important or are spam.
Regression	It is a supervised learning technique used to predict a continuous, numeric output value based on one or more input variables. It is used to model and understand the relationship between the input features and the target variable.	Using past sales data to predict future sales.
Clustering	It is an unsupervised learning technique used to group similar data points together based on their inherent characteristics. It is used to identify patterns and discover underlying structures within the data without any predetermined classes.	Grouping similar customer profiles based on their purchase history.

[198] De Silva, D and Alahakoon, D. (2022). An Artificial Intelligence Life Cycle: From Conception to Production. *Patterns, 3*, 4-6.
[199] Laato, S et al. (2022). AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering. *Proceedings - 1st International Conference on AI Engineering - Software Engineering for AI, CAIN*, 115. .

Anomaly Detection	It is the identification of data points, events, or observations that deviate significantly from the normal or expected behaviour within a dataset. It is used to identify rare, unusual, or suspicious occurrences that may require further investigation.	Prevention of money laundering or fraudulent transactions.
Recommendation	It is a technique used to suggest relevant items, products, or content to users based on their preferences, behaviours, and the preferences of similar users.	Internet marketing and shopping suggestions.
Object Detection	It is a computer vision technique used to identify and locate specific objects within an image or video frame. It is used to recognize and classify objects, people, or other elements of interest within a given scene.	Systems for traffic control that can differentiate between cars, motorbikes, trucks, and give real-time statistics on the number of vehicles on a road.
Prediction	It is the use of historical data and patterns to forecast or estimate future events, outcomes, or behaviours. It involves the application of various machine learning and statistical techniques to make informed predictions about future occurrences.	Predict crime rates in cities (indicating hot spots where police efforts can be focused at certain times)

Figure 6. Most common functionalities of artificial intelligence. By Lorena Flórez.

Finally, from a governance perspective, the identification of the key AI functionalities for each project can help to fulfil not only the Fundamental Rights and Algorithm Impact Assessment (FRAIAs) to understand how the AI system works, but also other impact assessments according to the industry and the functionality of the model (e.g. Privacy (PIA), Data Protection (DPIA), Technology (TIA), Environmental (EIA)). Beyond the technical aspects, the implementation stage also involves addressing the governance and ethical considerations surrounding the use of AI systems (*governance of AI*).[200] This includes establishing clear policies and protocols for the use of AI, ensuring transparency and accountability in decision-making processes, and addressing potential biases and discrimination that may arise from the AI models. The life cycle of an AI system is an iterative process, requiring continuous monitoring, evaluation, and refinement to ensure the responsible and effective deployment of these technologies.

[200] Flórez Rojas, M.L. (2023). Pensamiento de Diseño y Marcos Éticos Para La Inteligencia Artificial: Una Mirada a La Participación de Las Múltiples Partes Interesadas. *Desafíos, 35(268)*, 344-345.

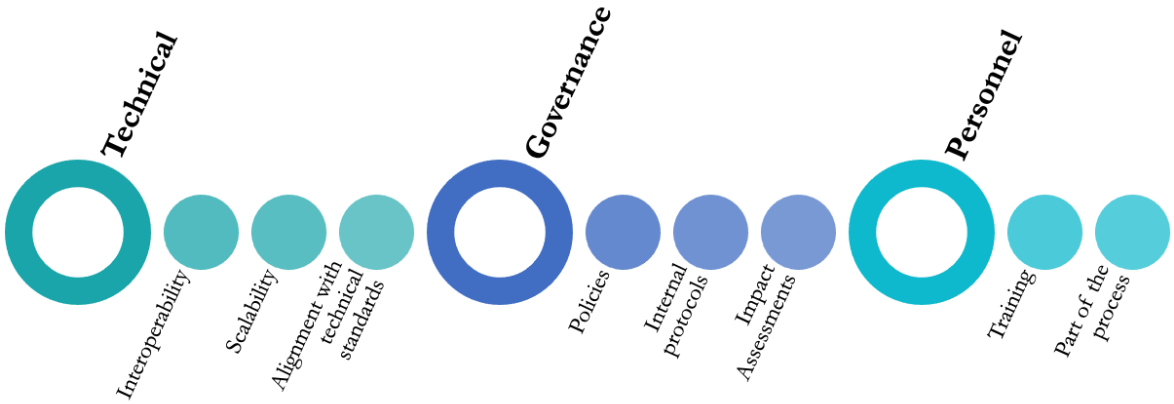


Figure 7. Elements of the organisational implementation of AI systems stage. By Lorena Flórez.

3.3 Fundamental rights issues in the implementation stage- Study cases of real applications

In the following section, we explore the implementation and organisational embedding of AI systems through the lens of real-world case studies and applications. Considering the previous elements of technological integration, system configuration, and governance challenges, we analyse how these issues have manifested in the practical deployment of AI tools in the public sector. These case studies highlight the diverse ways in which organisations have navigated the integration of AI systems into their existing operations, as well as the strategies employed to address concerns around transparency, accountability, and the mitigation of potential biases. Furthermore, where real-world examples may be limited or emerging technologies are involved, we will also analyse possible implications for fundamental rights in a hypothetical manner. This approach allows us to anticipate potential challenges and consider preemptive measures, even in scenarios that have not yet fully materialised in practice.

Austria: Email management, anonymization and file management

Austria has implemented the use of AI-powered tools for various document management tasks within its court system.[201] One of the primary applications is the **anonymization of court decisions.** The AI system is used to recognize and prepare proposals for the anonymization of names, addresses, and roles of parties involved in judgments and decisions. This helps to prepare the court decisions for publication while protecting the privacy of the individuals mentioned.[202] Additionally, the AI system is used as a **digitization assistant for**

[201] Kalliopi, T. (2023). Automated Anonymization of Court Decisions, ACM. Nineteenth International Conference on Artificial Intelligence and Law, *5, 5-7*.

[202] Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung. (2022). Einsatz von künstlicher Intelligenz bei der Anonymisierung von Gerichtsentscheidungen: Justiz gewinnt eAward 2022. <https://www.brz.gv.at/presse/eaward2022.html.> See also, Hackl, M. and Steinbauer, D. (2021) Anonymization of Court Decisions. <https://commission.europa.eu/document/download/3555369b-1c60-4e09-acb5-5e578cb3c2e6_en?filename=anonymisation_webinar_29032021_ms_austria>.

analogue court files. The system can determine the description and classification of scanned documents, detect the first pages of documents and attachments, and propose a digital file structure for the scanned materials. This helps to streamline the process of digitising the court's analogue records. The AI system also plays a role in **managing the intake of digital and physical legal communications**. It can analyse, classify, and extract metadata from incoming electronic legal communications and scanned documents.[203] The system then routes these documents and creates files, while also proposing naming conventions.

However, the anonymization of court decisions remains a largely manual task, despite the AI system providing supporting functionality. This is due to insufficient human resources at the courts, resulting in the Supreme Court being the primary publisher of anonymized decisions.[204] To address this challenge, the courts are exploring the use of multiple AI technologies to further automate the anonymization process. This includes pre-analysing the database of anonymized and non-anonymized decisions, selecting appropriate methods and algorithms, and designing search-based and natural language processing models to enhance the automation capabilities.

The implementation of AI in email handling and document anonymization presents both opportunities and challenges, particularly in **multilingual and multicultural contexts.**[205] In countries with diverse populations, AI systems may encounter difficulties in accurately anonymizing names that do not conform to the country's typical naming conventions. For instance, names or surnames in different languages or scripts may not be properly recognized or anonymized by the AI, potentially compromising **privacy.** This issue highlights the necessity of human oversight in the anonymization process.

Furthermore, automated email distribution systems could **inadvertently hamper access to justice** if important court communications are incorrectly flagged as spam or routed to the wrong recipients. To address these concerns, constant auditing and revision of automated systems are essential to protect fundamental rights and ensure equal access to justice for all citizens, regardless of their linguistic or cultural background. The **human-in-the-loop** approach becomes crucial in these scenarios, allowing for manual intervention and correction of AI-generated errors, especially in cases involving complex or atypical names and multilingual documents.

---

[203] Stawa, G. (2018). Artificial Intelligence: How is Austria approaching AI integration into judicial policies. Federal Ministry of Constitutional Affairs, Reforms, Deregulation and Justice Wie. <https://rm.coe.int/how-is-austria-approaching-ai-integration-into-judicial-policies-/16808e4d81>.

[204] Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung. (2022) Mein Name ist Nym, Ano Nym. Bundesrechenzentrum. <https://www.brz.gv.at/was-wir-tun/Innovationen/Mein-Nameist-Nym--Ano-Nym.html>.

[205] Nuno, M. et al. (2023). Hallucinations in Large Multilingual Translation Models. Transactions of the Association for Computational Linguistics. *11. 1505-1507.*

Therefore, this case highlights the potential benefits of integrating AI systems into court operations, but also underscores the need for careful implementation, human oversight, and continuous refinement to ensure the responsible and effective deployment of these technologies. These applications highlight the potential benefits of integrating AI systems to streamline administrative processes and enhance efficiency within the judicial sector. However, it also underscores the importance of careful implementation and organisational embedding of such technologies. For instance, the **initial anonymization of court decisions** was still a largely manual process, with the AI system providing supporting functionality rather than fully automating the task. This reflects the need to gradually introduce and test AI-powered tools, **ensuring appropriate human oversight** and the ability to **refine the system** based on user feedback. The project's approach of pre-analysing the database, selecting appropriate methods and algorithms, and combining various AI techniques, such as natural language processing (NLP) and machine learning (ML), demonstrates the technical complexities involved in configuring the system to meet the specific requirements of the courts.

Belgium: AI platform for public decision-making insights

GoVocal, formerly CitizenLab, is a civic technology company in Belgium, developed a public participation platform that uses machine learning (ML) algorithms to help public officials process and analyse large volumes of citizen input efficiently.[206] The platform employs Natural Language Processing (NLP) and machine learning techniques to automatically classify and analyse thousands of contributions collected from citizen participation platforms.

A notable example of its application was the Youth4Climate platform, created in response to youth protests against climate change inaction. The AI system analysed 1,700 ideas, 2,600 comments, and 32,000 votes submitted by users, identifying and grouping the most important and supported priorities.[207] The AI-driven findings were used to develop a report with 16 policy recommendations for elected officials. Therefore, cities using this platform, such as Kortrijk and Temse, have reported positive outcomes. The platform has helped them process citizen contributions more easily, group ideas into main themes, and make informed decisions about resource allocation based on citizen input.[208]

The CitizenLab case illustrates several key aspects of the implementation stage of the AI lifecycle. First, it demonstrates the importance of integrating AI systems into existing workflows and processes (*integration and interoperability*). The platform was designed to

[206] https://www.govocal.com/.

[207] Sgueo, G. (2020). The Practice of Democracy A Selection of Civic Engagement Initiatives. *26*. <http://www.europarl.europa.eu/thinktank> See also Karhu, M., Häkkilä J. and Timonen, E. (2019). Collaborative Media as a Platform for Community Powered Ecological Sustainability Transformations - A Case Study, *ACM International Conference Proceeding Series, 4*. <https://dl.acm.org/doi/10.1145/3365610.3368413>.

[208] Van Aeken, K. (2017). Digital Democracy in Belgium and the Netherlands. A Socio-Legal Analysis of Technologies, Embedding and Expectations of Two Fourth Wave Innovations, Digital Democracy in a Globalized World, *Elsevier BV, 220-225*.

complement and enhance the public participation process, not to replace human decision-making entirely. Secondly, the case highlights the need for continuous iteration and refinement of AI systems. CitizenLab works to e**nsure optimal use of their automated dashboards by governments** and is exploring ways to apply the technology to larger-scale conversations on social media and other online forums.

The challenges faced by CitizenLab, particularly in terms of algorithm classification and human adoption, underscore the importance of scalability and adaptability in AI implementation. The need to tailor the system to different administrative workflows and taxonomies across regions demonstrates the complexity of deploying AI solutions in diverse contexts and languages. Finally, the case emphasises the critical role of **user trust** and understanding in successful AI implementation. CitizenLab learned that guiding users through the benefits of the technology and helping them interpret and trust the machine-generated results were crucial for adoption. This underscores the importance of **transparency, explainability, and effective human-AI interaction** in the implementation stage of AI systems in public administration.

3.4 Legal, regulatory and ethical frameworks applicable to implementation of AI systems used in public administration

The integration of AI into the public sector has the potential to significantly enhance operational efficiency, resource allocation, and decision-making processes, thereby reducing the workload of civil servants and potentially benefiting society at large. However, the effective implementation of AI systems within public organisational frameworks often faces substantial challenges, which can negatively impact the accountability and transparency of public administration. Three primary areas of concern can be identified in the implementation of AI systems within public administrative frameworks:

System Design: The autonomous learning capabilities of machine learning technologies, while advantageous, can minimise human supervision. This reduction in oversight may have disruptive effects on fundamental rights, as public officials often possess limited knowledge of system operations and minimal control over the decision-making process. Consequently, AI systems may increase legal uncertainty due to potential lack of accountability and transparency.[209]

Data Processing Framework: The efficacy of AI systems is heavily dependent on the availability and quality of data. Public administrations may face challenges in accessing "good data" due to restrictions on sensitive information, such as biometric data. Furthermore, there may be insufficient data available or a lack of an effective legal framework for data collection and protection.[210]

---

[209] Selten, F. and Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly, 41(1),* 1-14..

[210] Van Noordt, C. et. al. (2023). Policy initiatives for Artificial Intelligence-enabled government: An analysis of national strategies in Europe. *Sage Journals*, 1-39. .

Ethical Considerations: The ethical implications of AI systems in public administration are significant. The values embedded during the training process and the system's capacity to act in the public interest may be unclear or problematic.[211]

The primary concerns emerging from the deployment of AI systems by public authorities relate to algorithmic bias and data protection. The European Union has been at the forefront of addressing these issues through various policy proposals, most notably the AI Act, which aims to encourage innovation while safeguarding fundamental rights.
This section will proceed first with an analysis of legislative proposals at the EU level to regulate AI. Then, an examination of case law developed at both EU and national levels, demonstrating the challenges of effective implementation in the public sector. Finally, a description of relevant national strategies.

It is important to note that the legal framework governing the organisational implementation of AI and its implications for fundamental rights in public administration is still evolving at international, EU, and national levels. The following paragraphs will introduce key legal provisions concerning the organisational implementation of AI systems and the fundamental rights issues arising therefrom in the context of public administration. Additionally, we will present relevant case law from the Court of Justice of the European Union and national courts to illustrate the current legal landscape and emerging challenges.

3.5 European legal framework and case law

3.5.1 Legal framework

Council of Europe, Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data

The "Decision of the Committee of Ministers" from the 128th session held in Elsinore on May 18, 2018, represents a significant milestone in the modernization of data protection within the Council of Europe framework[212]. The Committee formally acknowledged Parliamentary Assembly Opinion No. 296 (2017) and adopted a Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108). This amendment aims to strengthen the application of the Convention and its Additional Protocol concerning supervisory authorities and transborder data flows (ETS No. 181).

[211] European Parliament. (2020) Artificial Intelligence: Threats and opportunities. <https://www.europarl.europa.eu/topics/en/article/20200918STO87404/artificial-intelligence-threats-and-opportunities>.
[212] Council of Europe. Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Decision of the Committee of Ministers 128th session. (2018) <https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/LIBE/DV/2018/09-10/Convention_108_EN.pdf>.

Key provisions of the Protocol include a broad definition of "data processing," encompassing any operation on personal data, whether automated or manual, within structured data sets accessible by specific criteria. Article 9 of the amended Convention guarantees individuals the right not to be subject to decisions significantly affecting them based solely on automated processing, without consideration of their views. This protection is crucial, especially in contexts where algorithmic decisions could result in the denial of rights, social benefits, or credit capacity based purely on automated reasoning.

The Protocol emphasises the importance of individuals' rights to challenge purely automated decisions, allowing them to present their perspectives and contest the accuracy, relevance, or fairness of the data and profiling used in such decisions. However, this right is limited in cases where the automated decision-making process is legally authorised and includes appropriate safeguards to protect the data subject's rights, freedoms, and legitimate interests. The modernization of ETS No. 108 through this Protocol thus reflects a commitment to adapting data protection laws to contemporary challenges posed by automated data processing, while ensuring that individuals' fundamental rights remain safeguarded.

Therefore, if any public organisation plans to implement AI systems or related technologies that could result in automated or semi-automated decision-making, the organisation must establish protocols and policies that allow citizens to challenge, be informed about, and provide feedback on these decisions. Similarly, the organisation should have an internal protocol to distinguish between systems that lead to fully automated decisions and those that do not, as well as to define the oversight procedures to ensure alignment during the implementation stage.

European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies

It outlines a comprehensive framework addressing the ethical aspects of AI, robotics, and related technologies. The resolution underscores the need for a harmonised regulatory framework grounded in Union law, the Charter of Fundamental Rights, and international human rights law, particularly for high-risk technologies.[213] This framework should ensure uniform standards across the EU, safeguarding fundamental values such as human dignity, autonomy, and transparency while promoting fairness and preventing bias and discrimination, particularly against minority groups.

The resolution advocates for a legal framework that includes both binding legal obligations and ethical principles to govern the development, deployment, and use of AI technologies. These principles should ensure that AI serves to enhance human well-being, remains transparent, and is explainable, while also addressing the risks associated with AI, such as

[213] European Parliament. (2020). Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020 P9_TA 0275) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020IP0275>. .

algorithmic bias and the ethical use of Big Data. Additionally, it calls for the integration of non-technical disciplines in AI research to ensure a comprehensive understanding of its societal impacts.

A significant emphasis is placed on a risk-based approach to regulation, where AI technologies that pose significant risks to individuals or society, particularly in terms of fundamental rights and safety, should be classified as high-risk and subjected to stringent oversight. This classification should be determined through an impartial, regulated ex-ante assessment based on well-defined criteria, including the sector of application and the specific use or purpose of the technology.

The resolution also highlights the importance of data quality, particularly the representativeness and de-biasing of training data sets and calls for these data sets to be auditable by national supervisory authorities. Governance measures that focus on accountability and address potential biases are seen as crucial for increasing public trust in AI technologies. Moreover, the resolution stresses that public procurement processes should adhere to the same ethical standards required of public authorities, even when AI development and deployment are outsourced to private entities.

If any public organisation plans to implement AI systems or related technologies, it must establish appropriate technical ex-ante measures to fully understand and agree on the type of system being implemented, as well as its various functionalities, including its limitations. Additionally, the organisation should have an internal protocol to assess the quality of the data, ensuring that data processing principles are effectively applied.

Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence

It addresses the civil liability regime for AI, acknowledging the significant legal challenges posed by certain AI systems under existing liability frameworks. The resolution recognizes that AI's complexity, opacity, and potential autonomy can make it difficult to identify the specific cause of harm or damage, particularly when the responsible code, input, or data is obscured.[214] This opacity risks leaving victims without adequate compensation, as it complicates the process of linking harm to its cause.

Despite the need for adjustments, the resolution suggests that a complete overhaul of current liability regimes is unnecessary. Instead, it recommends **specific and coordinated modifications** to address the unique challenges AI systems present, such as their ability to self-learn, their vulnerability, and the involvement of multiple actors in their development and deployment. The resolution emphasises that while AI systems can cause harm directly or indirectly, legal liability should remain with the human actors—those who build, deploy, or

[214] European Parliament. (2020). Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) P9_TA(2020)0276. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020IP0276>.

interfere with these systems—rather than granting legal personality to the AI systems themselves. By holding all parties in the AI value chain accountable, the resolution aims to ensure that victims receive appropriate compensation and that the existing liability frameworks remain effective in the face of emerging AI technologies.

The Artificial Intelligence Act (AIA)

It is the world's first comprehensive regulation on artificial intelligence, officially entering into force on August 1st, 2024, marking a significant step toward ensuring that AI developed and used in the EU is trustworthy and aligned with fundamental rights.[215] The AI Act aims to establish a harmonised internal market for AI within the Union, fostering innovation and investment while safeguarding human rights.

The AI Act introduces a forward-looking definition of AI, rooted in a product safety and risk-based approach.[216] AI systems are categorised into four risk levels:

Risk level	Description	Example
Minimal	Most AI systems, such as AI-enabled recommender systems and spam filters, fall into this category. These systems are subject to minimal or no obligations under the AI Act due to their low risk to citizens' rights and safety[217]. Organisations may choose to voluntarily adopt additional codes of conduct.	AI system that automatically filters all incoming emails and tells us which ones are not so important or are spam:
Recommendations: Public organisations using this AI system should clearly notify citizens that such a system is in place. This could be done by including a notice on relevant public platforms, such as the official website or public registry of the country. The notice should explain the purpose of the system, how it works, and the criteria it uses to filter emails. A. Explanatory Documentation: To further ensure transparency, the organisation should provide a one-page explanatory document that outlines the key features of the AI system. This document should include: • A brief overview of how the AI system filters emails. • Information on how citizens' emails are processed and classified. • Assurance that no personal data is stored or used beyond the purpose of filtering.		

[215] 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 (Artificial Intelligence Act).

[216] Malgieri, G. and Pasquale, F. (2024). Licensing High-Risk Artificial Intelligence: Toward Ex Ante Justification for a Disruptive Technology. Computer Law & Security Review. *52*, 1-18. See also, Schwemer, S. et al. (2021). Legal AI Systems in the EU's proposed Artificial Intelligence Act. Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace, held in conjunction with ICAIL 2. <https://ssrn.com/abstract=3871099>.

[217] Regulation (EU) 2024/1689. Recital 53, 165.

- Clear instructions on how individuals can challenge or inquire about the classification of their emails if they believe a mistake has been made.

User Awareness and Control: The organisation should also consider providing citizens with the option to opt-out of having their emails filtered by the AI system, or at least offer a way for them to review the filtering decisions. This empowers citizens and builds trust in the system by ensuring they retain some level of control over how their communications are handled.

Specific Transparency risk[218]	AI systems like chatbots and certain AI-generated content (e.g., deep fakes) must clearly disclose their nature to users. AI-generated content, including synthetic audio, video, text, and images, must be labelled and detectable as artificially generated.[219] Providers are required to design systems that ensure such content is marked in a machine-readable format.	A public administration uses an AI-powered chatbot on its official website to assist citizens with inquiries and provide information.

Recommendations: Public organisations using this AI system should clearly notify citizens that such a system is in place.

A. Disclosure: The chatbot must clearly inform users at the start of each interaction that they are communicating with an AI system. For example, a message like "*You are now chatting with an AI assistant. How can I help you today?*" should be displayed.[220]

B. Labelling of AI-Generated Content: Any content generated by the chatbot, such as responses, recommendations, or documents, must be clearly marked as AI-generated. For instance, at the bottom of each response, a note could say, "*This response was generated by an AI system.*"

C. Machine-Readable Marking: The system should be designed to ensure that all AI-generated content, whether text, audio, or video, is marked in a machine-readable format. This allows other systems and tools to recognize the content as AI-generated, ensuring transparency and preventing misuse.

High Risk	High-risk AI systems are those that significantly impact safety or fundamental rights.[221] These systems must comply with strict requirements, including risk mitigation, high-quality data sets, logging of activity, detailed documentation, clear user	An AI-powered criminal recidivism prediction system used by courts to inform sentencing and parole decisions. This system would be considered high-risk because:

[218] The AI Act also introduces specific rules for general-purpose AI models, which are highly capable and designed to perform a wide variety of tasks, such as generating human-like text. The Act ensures transparency across the value chain and addresses potential systemic risks associated with these models.

[219] Regulation (EU) 2024/1689. Article 40, Chapter V. Seccion 1.

[220] GPAI. (2023). Scaling Responsible AI Solutions: Learning from AI teams to identify and address challenges in Responsible AI. Report. *Global Partnership on AI.* 23-25. <https://gpai.ai/projects/responsible-ai/scaling-responsible-ai-solutions/>.

[221] Regulation (EU) 2024/1689. Chapter III.

	information, human oversight, and robust cybersecurity measures.[222] Regulatory sandboxes will facilitate the responsible innovation and development of compliant AI systems.[223] High-risk AI systems include those used for recruitment, credit assessments, and autonomous robots.	● It significantly impacts fundamental rights (liberty, fair trial). ● It is used in the administration of justice and democratic processes. ● It has the potential for bias and discrimination.

Recommendations: Public organisations using this AI system should implement more robust technical and legal measures:

- Conduct thorough risk assessments: Identify potential risks to fundamental rights and safety. Evaluate the system's impact on vulnerable groups (e.g. take into account the country's diverse groups)
- Ensure high-quality datasets: Use diverse, representative, and up-to-date data. Regularly check for and mitigate biases in training data.
- Implement robust documentation practices: Maintain detailed records of the system's development, testing, and deployment. Document decision-making processes and system limitations.
- Establish human oversight: Ensure human review of AI-generated recommendations. Implement a "human-in-the-loop" approach for critical decisions.
- Provide clear information to users: Explain how the AI system works and its limitations. Offer transparency about when and how AI is being used.
- Implement strong cybersecurity measures: Protect the system from unauthorised access and data breaches. Regularly update and patch the system.
- Ensure logging and traceability: Implement comprehensive logging of the system's activities. Ensure decisions can be traced back to specific inputs and processes
- Participate in regulatory sandboxes: Engage with regulators to test and refine the system in controlled environments. Use feedback to improve compliance and performance.
- Regular auditing and monitoring: Conduct periodic audits to ensure ongoing compliance. Monitor system performance and outcomes for unexpected behaviours or biases.
- Provide training and support: Offer comprehensive training to staff using the AI system. Establish support mechanisms for addressing issues or concerns.

Unacceptable	AI systems that pose a clear threat to fundamental rights are banned under the Act. This includes systems that manipulate human behaviour, such as toys with voice assistance encouraging dangerous behaviour in minors, systems that enable social scoring by governments or companies, and certain predictive policing applications.[224] Additionally, some uses of biometric systems are prohibited, such as emotion recognition in workplaces and real-time	A government-developed social scoring system that rates citizens based on their online behaviour, social media posts, and public interactions to determine their access to public services and job opportunities.

[222] Regulation (EU) 2024/1689. Article 9.
[223] Regulation (EU) 2024/1689. Chapter VI.
[224] Chapter II. Article 5 AIA.

	remote biometric identification for law enforcement in public spaces, with narrow exceptions.

Recommendations: This system might fall under the unacceptable risk category. Therefore, it would be best to focus financial resources and efforts on developing assistance tools for public bodies, rather than on categorising citizens in ways that might affect human rights.

Figure 7. The AI Act's risk-based approach explained with examples. By Lorena Flórez.[225]

In this regard, and considering the implementation phase of AI in a public organisation, it is important to classify and support the classification of the systems being implemented or planned for implementation. The organisation must ensure that each system is **adequately classified** and that all obligations are fulfilled. For instance, if the organisation is implementing an AI system to classify emails, it should meet the necessary transparency requirements to assure the community that the labels applied do not pose high risks to society. On the other hand, if the public body intends to implement a system that gathers information from citizens, as in the case of the CitizenLab example, there is a potential risk of overshadowing certain opinions or introducing bias depending on the labels used. This could threaten fundamental rights such as freedom of expression and participation in public debate. In such cases, the public body should not only provide transparency regarding access to information about the system but also offer clear and simple explanations of the decision-making process. This would include how the system collects public opinions and comments, the labels chosen by the organisation, and the reasons behind those choices.

Special Report. EU Artificial intelligence ambition: Stronger governance and increased, more focused investment essential going forward

The European Union's pursuit of leadership in AI has been marked by ambitious plans and significant challenges, as outlined in the special report on EU AI ambition.[226] The report critically examines the effectiveness of the Commission's contribution to developing the EU's AI ecosystem, focusing on the coordination of measures outlined in the 2018 and 2021 AI plans, the adoption of a legal framework for data sharing and trustworthy AI, and the implementation of EU-funded infrastructure and research initiatives. The audit reveals a complex landscape where aspirations for AI excellence and trust intersect with practical hurdles in governance, investment, and implementation.

The report identifies several key areas of concern. Firstly, the coordination between Commission and national measures was found to be ineffective due to limited governance tools and outdated targets. EU AI investment lagged behind global competitors, and the

[225] **Disclaimer:** The examples provided in the table are not to be taken as legal consultancy. The applicability of these examples can vary depending on system functionalities, country, and other specific situations. They are intended solely for illustration and general exemplification of the risk-based approach.

[226] European Court of Auditors. (2024). Special Report. EU Artificial intelligence ambition: Stronger governance and increased, more focused investment essential going forward. (08/2024). <https://www.eca.europa.eu/ECAPublications/SR-2024-08/SR-2024-08_EN.pdf>.

implementation of infrastructure and capital support for SMEs was slower than anticipated. While the Commission successfully increased EU budget spending on AI research, the monitoring of these projects' contributions to the EU AI ecosystem was inadequate. The EU's AI plans, though comprehensive compared to international counterparts, lacked specific performance targets and a robust monitoring framework. The report also highlights the critical role of national involvement in mobilising AI investment, though the member states' contributions to overall EU investment targets remained unclear.

In response to these findings, the report offers several recommendations for the Commission. These include reassessing the EU investment target for AI and agreeing with member states on their contributions, evaluating the need for an EU-funded capital support instrument for AI-innovative SMEs, ensuring coordinated operation of EU-funded AI infrastructure, and improving the tagging, targeting, and monitoring of AI-related research and innovation spending across the EU budget. The report makes a call to review AI ecosystems (i.e. systems of interdependencies between public and private players involved in the research, innovation, production and consumption of AI) are key to fostering research and innovation in this field. Therefore, public administration bodies should also explore different ways of seeking partnerships with private and other public agencies to collaborate into creating a better AI ecosystem.

3.5.2 Case law

In the following section, we explore the implementation and organisational embedding of AI systems through the lens of some cases and applications. Considering the previous elements of technological integration, system configuration, and governance challenges, we analyse how these issues have manifested in the practical deployment of AI tools. These cases highlight the main elements and features that have been addressed in a court of law considering the implementation of an AI system.

Germany: Bundesverfassungsgericht- The Gotham system

In 2017, the federal region of Hesse introduced the use of the Gotham system, an AI platform developed by the American company Palantir, for crime prevention purposes.[227] The software enables predictive policing by allowing investigators to quickly gather and analyse information about potentially dangerous individuals and situations. This is achieved by drawing connections between people, places, and objects using data from multiple sources. The system was intended to enhance crime prevention efforts and improve the efficiency of police responses.[228]

[227] BVerfG, Judgment of the First Senate of 16 February 2023 - 1 BvR 1547/19 -, paras. 1-178. Press Release No. 18/2023 of 16 February 2023.

[228] For more information about the system background and application in Germany see Ulbricht, L. and Egbert, S. (2024). In Palantir We Trust? Regulation of Data Analysis Platforms in Public Security. *Big Data and Society. 11*. 4-8. .

However, on February 16, 2023, the Federal Constitutional Court (Bundesverfassungsgericht) ruled that certain provisions of the Hesse (HSOG) and Hamburg (Hmb-PolDVG) laws, which permitted the processing of stored data through automated analysis (Hesse) and interpretation (Hamburg), were unconstitutional. The court found that these laws **violated the right to informational self-determination** as recognized by the Basic Law (Grundgesetz). The court's primary concerns were the overly broad powers granted to authorities and the lack of limitations on the volume of data and the methods of processing. The court emphasised that **automated methods can significantly interfere** with the right to informational self-determination and that such interference can only be justified when protecting legal interests of substantial weight, such as life, limb, or liberty. Additionally, less severe interferences might be justifiable for crime prevention, but only if there is a clear and identifiable danger. The court concluded that the required threshold for interference had not been met.

This case underscores the challenges of **accountability, transparency, and potential bias** in AI-driven data processing. The methods of data collection were not sufficiently transparent, leaving the public unclear about how their data was being used and processed. This lack of transparency can lead to reduced accountability for public authorities and limits opportunities for independent assessments of the system's effectiveness. Moreover, the case highlights that **not only the rights of suspects were at risk, but also those of individuals with no connection to the suspects**, due to the vague legislative guidelines on the type of data that could be used. The expansive data collection methods, such as data mining, pose significant privacy and fundamental rights concerns, particularly when machines predict potentially dangerous individuals before any crime has been committed. This case highlights the crucial need for public authorities to be transparent about their use of AI systems and the specific purposes for which these systems are employed.

The challenges identified in this case serve as a critical reminder for public administrations considering the implementation of AI systems. It is essential that during the implementation stage, public bodies establish clear protocols and limitations on data collection and processing. They must also ensure transparency, enabling citizens to understand how AI systems are used and for what purposes, and ensuring that their rights are protected. Proper oversight and accountability mechanisms must be in place to prevent misuse and to uphold public trust in these technologies.

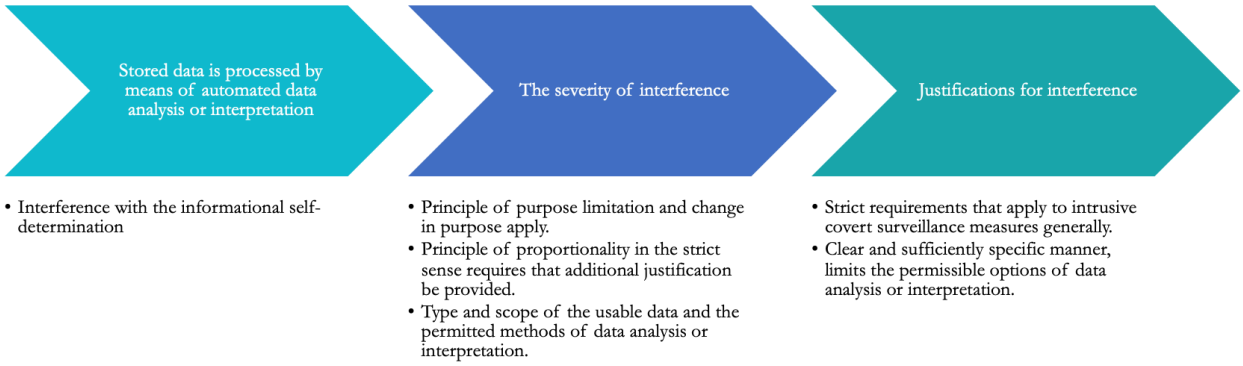


Figure 8. The processing of stored data through automated analysis and interpretation can be unconstitutional. By Lorena Flórez based on the Gotham system.[229]

Poland: Ministry of Labour and profiling of unemployed people

In May 2014, the Polish Ministry of Labour introduced an Automated Decision-Making (ADM) system designed to address unemployment by categorising job seekers into three distinct groups based on their skills.[230] The categorization process began with a computer-based interview, followed by the input of various themes into a database, each of which was assigned a specific score. The system considered two key variables: "distance from the labour market" and "readiness to enter or return to the labour market." The first criterion assessed factors that hinder market entry, such as gender, age, and educational background. The second criterion evaluated the individual's motivation for seeking employment. Ultimately, an algorithm determined the final score and assigned the individual to one of the three categories.[231]

The ADM system faced significant criticism due to its perceived **lack of transparency, data protection issues, and discriminatory effects** during its implementation. One of the main transparency concerns was that individuals were unaware of how much weight each question carried in determining their final categorization.[232] This opacity not only impacted the rights of those using the platform but also posed challenges for data subjects seeking to enforce

[229] Analysis based on the judgement BVerfG, Judgment of the First Senate of 16 February 2023 - 1 BvR 1547/19 -, paras. 1-178. Press Release No. 18/2023 of 16 February 2023.

[230] MLSP Ministry of Labor and Social Policy. (2014). Profilowanie pomocy dla osób bezrobotnych. Podręcznik dla pracowników powiatowych urzędów pracy. <https://panoptykon.org/sites/default/files/podrecznik_profilowania.pdf>. .

[231] Jędrzej Niklas et. al. (2015). Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making. 18-20. <https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf>. .

[232] Ibid. 33. See also Sztandar-Sztanderska, K. and Zieleńska, M. (2022). When a Human Says "No" to a Computer: Frontline Oversight of the Profiling Algorithm in Public Employment Services in Poland. *Sozialer Fortschritt. 71(*6), 468.

their rights under the GDPR.[233] According to Article 15(1) GDPR, data subjects have the right to access their data and understand the purposes for which it is used and by whom. Additionally, Article 16 GDPR grants them the right to rectify incorrect or incomplete data. However, the lack of transparency in the system made it difficult for individuals to access, review, or correct their responses.

Moreover, the unclear decision-making process raised concerns regarding Article 22(1) GDPR, which grants individuals the right not to be subject to decisions based solely on automated processing.[234] It was uncertain whether the final categorization was entirely automated or whether human intervention occurred at any stage. The Supreme Audit Office (Najwyższa Izba Kontroli) assessed the system and found that it categorised individuals differently based on gender, physical ability, and residential area, leading to potential violations of non-discrimination principles outlined in Article 21 of the Charter of Fundamental Rights (CFR) and Directive 2000/78/EC.[235] The system's rigid categorization also marginalised those in Profile III, often individuals with disabilities, by allocating fewer resources and opportunities to them.[236] Finally, in 2018, the Polish Constitutional Court ruled the system unconstitutional due to its lack of clarity in data collection principles and procedures.[237] However, the Court controversially upheld the provision that prevents individuals from challenging the system's decisions.

The Polish ADM system case underscores the critical importance of establishing clear governance and technical requirements when implementing AI systems in public administration. To align better with legal transparency obligations, public bodies must ensure that AI systems provide clear, understandable information to citizens about how decisions are made, including the specific weight of each factor in the decision-making process. Additionally, policies and protocols should be in place to disclose whether the system operates fully automatically or with human oversight, highlighting the main risks to citizens.

Technically, public authorities must fine-tune such systems to prevent discriminatory outcomes, as seen in the system's biassed categorization based on gender and physical ability. Effective governance structures should be implemented to allow citizens to seek redress, ensuring that their rights under the GDPR are protected. By addressing these issues during the implementation stage, public administrations can mitigate the risks associated with AI systems and foster greater public trust.

---

[233] Ibid. 20.
[234] Ibid. 28.
[235] Najwyższa Izba Kontroli. (2019). Efektywność świadczenia usług rynku pracy. Informacja o wynikach kontroli, Delegatura w Lublinie. <https://www-nik-gov-pl.translate.goog/kontrole/P/18/079/LLU/?_x_tr_sl=pl&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=sc>. .
[236] Sztandar-Sztanderska, K. and Zieleńska, M. (2022). When a Human Says "No" to a Computer: Frontline Oversight of the Profiling Algorithm in Public Employment Services in Poland. *Sozialer Fortschritt. 71(*6), 475.
[237] Polish Constitutional Court. K-53/16. <https://ipo.trybunal.gov.pl/ipo/Sprawa?cid=1&dokument=16959&sprawa=18355>. .

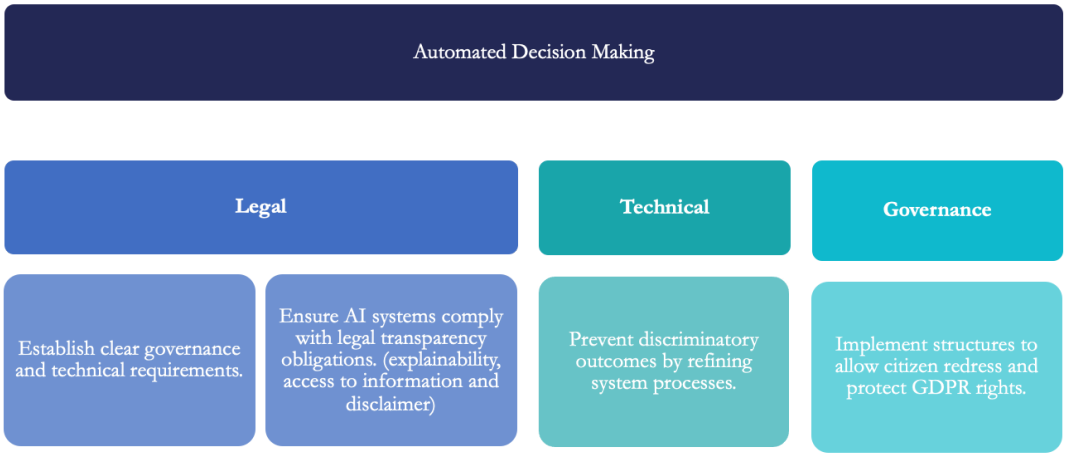


Figure 9. Main recommendations for ADM system. By Lorena Flórez based on the Polish profiling of the unemployed system.

France: The case of the Parcoursup Algorithm

The Parcoursup is the national pre-registration platform for access to introductory courses in undergraduate university education.[238] The algorithm was developed to assist admission offices in selecting candidates for higher education programs in France. Its implementation sparked a legal challenge before the Constitutional Council, primarily based on concerns about transparency in the decision-making process.

The main argument presented by applicants was that Parcoursup failed to provide sufficient information about how decisions were made, potentially violating the **right to access administrative documents** as guaranteed by Article 15 of the Declaration of the Rights of Man and Citizens. However, the Constitutional Council ruled that the Parcoursup algorithm did not violate this right.[239] The Council's decision was based on several key points:

- Restricting access to certain administrative documents was deemed necessary to protect the independence of teaching teams and maintain the secrecy of the deliberation process.[240]
- Candidates could access the general characteristics of the selection process before applying, as these were based on national standards.[241]
- Educational institutions were required to publish reports detailing the criteria used and the extent of algorithmic involvement, while respecting candidates' privacy.[242]

[238] Decree no. 2018-172 of 9 March 2018. <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000036692487>. .
[239] Council of State of France, ECLI:FR:CECHR:2019:427916.20190612, 12 June 2019, [para 18].
[240] Ibid. [para 13].
[241] Ibid. [para 15].
[242] Ibid. [para 17].

- In cases of adverse decisions, students could request specific information about the decision-making process, including the educational reasons and the weight given to different criteria.
- The decision-making process was not fully automated, as educational establishments set the conditions under which applications were considered, ensuring a degree of human intervention.[243]

This case demonstrates how AI systems can be made more accountable through human oversight and publicly available information about the decision-making process. Unlike the SyRI case, where data subjects were often unaware of algorithmic decision-making, Parcoursup applicants knew that an algorithmic system was involved in evaluating their applications.[244] This highlights the importance of disclosing the use of AI systems in public authority tasks.

However, the implementation phase revealed significant challenges in terms of **governance and accountability**. The lack of a clear governance structure with well-defined responsibilities for providing information to citizens has affected the transparency principle and accountability. It remains unclear who is responsible for providing information, to what extent this information should be publicly available, and on what grounds certain information may be withheld. This ambiguity in the implementation phase has hindered the full realisation of transparency and accountability principles, highlighting the need for more robust governance frameworks when deploying AI systems in public administration.

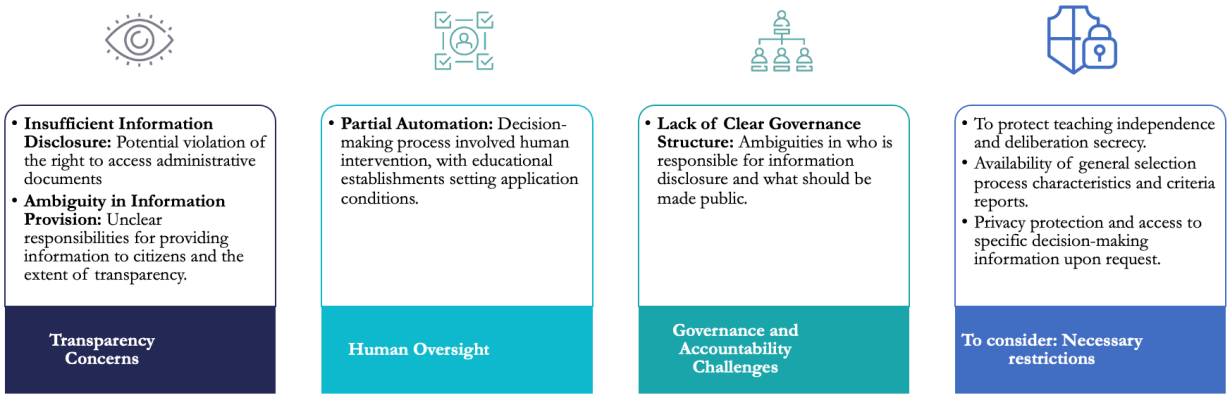


Figure 10. Main recommendations for ADM system. By Lorena Flórez based on the Parcoursup Algorithm.

---

[243] Ibid. [para 16].

[244] European Union Agency for Fundamental Rights. (2019). Fundamental Rights Report 2019. 160. <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-frr-chapter-7-infosoc_en.pdf> See also, Wolswinkel, Johan. (2022). Comparative Study on Administrative Law and the Use of Artificial Intelligence and Other Algorithmic Systems in Administrative Decision-Making in the Member States of the Council of Europe. 22. <https://www.coe.int/documents/22298481/0/CDCJ%282022%2931E+-+FINAL+6.pdf/4cb20e4b-3da9-d4d4-2da0-65c11cd16116?t=1670943260563>.

Denmark: The Gladsaxe model

In 2017, Gladsaxe municipality (in the suburbs of Copenhagen) had started to develop an algorithmic model trained to identify children with special needs from an early stage. The initial purpose was to identify at-risk children using a points-based analytical system.[245] It included factors like parental mental health, unemployment, missed medical appointments, and divorce. However, the automated model combined several risk indicators to identify 'ghetto' neighbourhoods which are then subjected to 'special measures' including higher punishments for crimes, removal of tenant protection to privatise public housing among others.[246] As part of the development process, the municipality asked for exemption from the data protection law to be able to combine personal (and sensitive) data from different data sources.[247]

By late 2018, the project was put on hold due to the model's relatively high error rate, which was partly a result of the limited historical data available—only 117 cases involving vulnerable children aged 0-6 years were used as training data for the model. Although the specific project was **never fully implemented**, it illustrates how the municipality envisioned using data-driven models to support vulnerable groups like children.[248] The model aimed to categorise families based on strong statistical correlations between selected data points. The intention was to use the model regularly and notify caseworkers when data flagged a specific family. Based on the flagged data, the caseworker would then decide if a specific intervention was warranted, with the family's consent required for further investigation.

However, it was addressed that the model posed discrimination risks based on ethnicity, race and migration. Later on the model faced heavy criticism for its approach, leading to its suspension.[249] Despite the hold, Gladsaxe municipality continued working on the algorithm, incorporating additional data sources. In 2019, after denial of permission by data protection authorities and negative press coverage, the project was stopped without further explanation. Nevertheless, in 2020, the University of Århus announced a new research project to develop an 'algorithmic tool for decision support' for detecting vulnerable children, which also faced criticism for similarities to the Gladsaxe model.

---

[245] Ministry of Social Affairs and the Interior. (2017). Dataunderstøttet tidlig opsporing af udsatte børn. <https://im.dk/nyheder/nyhedsarkiv/2017/dec/regeringen-siger-ja-til-33-nye-forsoeg-i-frikommunerne/>.

[246] Alfer, B. (2019) Automating society Denmark. *AlgoritmWatch*. <https://automatingsociety.algorithmwatch.org/report2020/denmark/>. .

[247] Gladsaxe Municipality. (2017). Ansøgning om konkrete forsøg i Frikommuneforsøg II. <https://sm.dk/media/6874/samlede-ansoegninger-frikommuneforsoeg-ii.pdf>.

[248] Jørgensen, R. (2023). Data and Rights in the Digital Welfare State: The Case of Denmark. *Information, Communication & Society, 26(123).*

[249] Kjær, J. (2018). Regeringen Vil Overvåge Alle Landets Børnefamilier Og Uddele Point. *Politiken* <https://politiken.dk/danmark/art6365403/Regeringen-vil-overv%C3%A5ge-alle-landets-b%C3%B8rnefamilier-og-uddele-point>. See also Alfter, B. (2020). Denmark - Automating Society Report 2020. <https://automatingsociety.algorithmwatch.org/report2020/denmark/>. .

This case serves as an example of the reliance on data analytics as a tool for governance, even in sensitive areas of public services. Unlike profit-driven predictions that target and influence consumer behaviour, these analytics are driven by the state's interest in predicting vulnerabilities among its citizens to direct specific interventions. However, this approach challenges citizen autonomy and agency in several ways. First, by collecting a large amount of personal data, including from citizens without special needs, to identify a relatively small group of at-risk families, the model interferes with the **right to privacy** of all citizens in the municipality. The approach taken by Gladsaxe municipality, however, appears to disproportionately favour the authorities' interest in experimenting with new data-driven analytics and predictions rather than adopting a human-centric approach that prioritises individual rights.[250] Second, automated decision support makes it more difficult for citizens to challenge the categorizations to which they are subjected and to contest being flagged as vulnerable (**effective remedy**).[251] Third, the notion that public interventions in such a sensitive area can be best guided by analysing historical data points, such as a family's medical appointments, reflects the institutional trust placed in data analytics to predict individual behaviour and guide public sector interventions.

Therefore this case underscores the ethical concerns and public resistance surrounding the use of AI profiling systems in sensitive areas like child welfare, even when intentions are ostensibly beneficial. It also highlights the importance of organisational embedding, not just from a technical perspective but also from a trust-building perspective. Implementing such a system should address the concerns of the end-users and consider their perspectives on the system's impact.

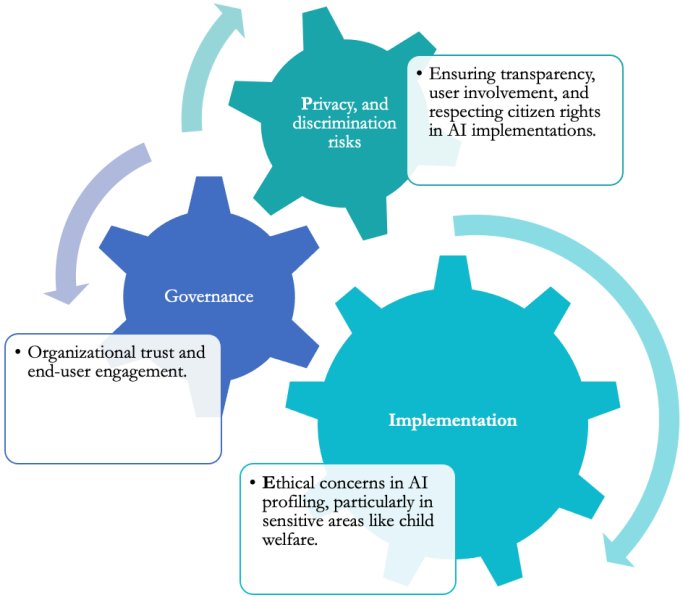


Figure 11. Implementation stage concerns. By Lorena Flórez based on the Gladsaxe model.

[250] Hoofnagle, C. J., Sloot, B. V. D., & Borgesius, F. Z. (2019). The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, *28*(1), 65–98.
[251] Bigo, D., Isin, E. F., & Evelyn, R. (2019). Data politics. In D. Bigo, E. F. Isin, & E. Ruppert (Eds.), *Data politics: Worlds, subjects, rights*.

Wales: Live Automated facial recognition technology: AFR Locate

The case of R (Bridges) v Chief Constable of South Wales Police concerned the legality of South Wales Police Force's (SWP) use of live automated facial recognition technology (AFR) in a trial system called *AFR Locate*. This system employed surveillance cameras to capture digital images of the public, which were then processed and compared with images of individuals on a law enforcement watch list.[252] The system was used openly, and the applicant's complaint was based on the grounds that AFR is incompatible with the right to respect for private life under Article 8 of the ECHR.[253]

The Court of Appeal Civil Division examined the legality of SWP's use of AFR Locate technology under the ECHR. In this regard, the Court concluded that the technology's implementation lacked a sufficient legal framework as required by Article 8 of the Convention.[254] Two primary flaws were identified in the framework regulating the implementation and use of AFR technology within organisational processes. The first flaw concerns the 'who' question: the criteria for including individuals on the watch list.[255] The second flaw relates to the 'where' question: the criteria for determining deployment locations.[256] Therefore, the Court found that **excessive discretion** was granted to individual police officers in both areas, with no clear guidelines on watch list inclusion or technology deployment locations.

The Court emphasised that AFR technology, which automatically captures and processes biometric data from many people, is often used in situations where most of the data is irrelevant to police investigations.[257] However, the UK legal framework did not specify clearly who may be included on watch lists or where the technology could be deployed. In addition, the issue of human intervention was also discussed. The lower divisional court had highlighted the presence of a 'human failsafe', requiring two persons to agree before police intervention based on a positive AFR match.[258] However, the Court of Appeal deemed this argument insufficient, noting that the Public Sector Equality Duty (PSED) refers to a '*duty as*

---

[252] EWCA Civ 1058, Case R Bridges v CC South Wales 2020 [para 1] See this report about the use of AFR in the United Kingdom Centre for Data Ethics and Innovation. (2020). Facial Recognition Technology. <https://www.gov.uk/government/publications/cdei-publishes-briefing-paper-on-facial-recognition-technology>.

[253] Ibid. [para 2].

[254] Ibid. [para 90-91]. See also how the use of AFR in Europe Neroni, I. (2020). Facial Recognition in Police Hands: Assessing the "Clearview Case" from a European Perspective. *New Journal of European Criminal Law 375 (11).*

[255] Ibid. [para 91].

[256] Ibid. [para 92]. See also the analysis of the case Kyriakos, N and Oswald, M. (2020). The Long Arm of the Algorithm? Automated Facial Recognition as Evidence and Trigger for Police Intervention. *Forensic Science International: Synergy 86 (2).*

[257] Marcus, M. and Smith,M. (2017). Automated Facial Recognition Technology: Recent Developments and Approaches to Oversight. *The University of New South Wales Law Journal 40(1).*

[258] Global Freedom of Expression. (2020). R v. the Chief Constable of South Wales Police. <https://globalfreedomofexpression.columbia.edu/cases/r-v-the-chief-constable-of-south-wales-police/>. .

to the process which needs to be followed, not what the substance of the decision should be'.[259] Thus, the Court determined that the DPIA failed to properly assess the risks to the rights and freedoms of data subjects and did not address the measures envisaged to address the risks arising from the identified deficiencies. This failure was seen as an inevitable consequence of the overly broad areas of discretion in selecting individuals for watch lists and deciding on deployment locations.[260]

In conclusion, the Court of Appeal deemed the legal framework, which the Divisional Court had considered sufficient to constitute the legal basis of AFR Locate for the purposes of Article 8(2), to be insufficient.[261] The Court decided that while the use of AFR technology engaged Article 8(1) ECHR, it was not in accordance with the law for the purposes of Article 8(2) due to the excessive discretion left to individual police officers and the lack of clear criteria for watch list inclusion and deployment locations.[262] This case highlights the importance of clear guidelines and limited discretion in the implementation of new surveillance technologies, particularly when they involve the processing of biometric data and potentially infringe on privacy rights.

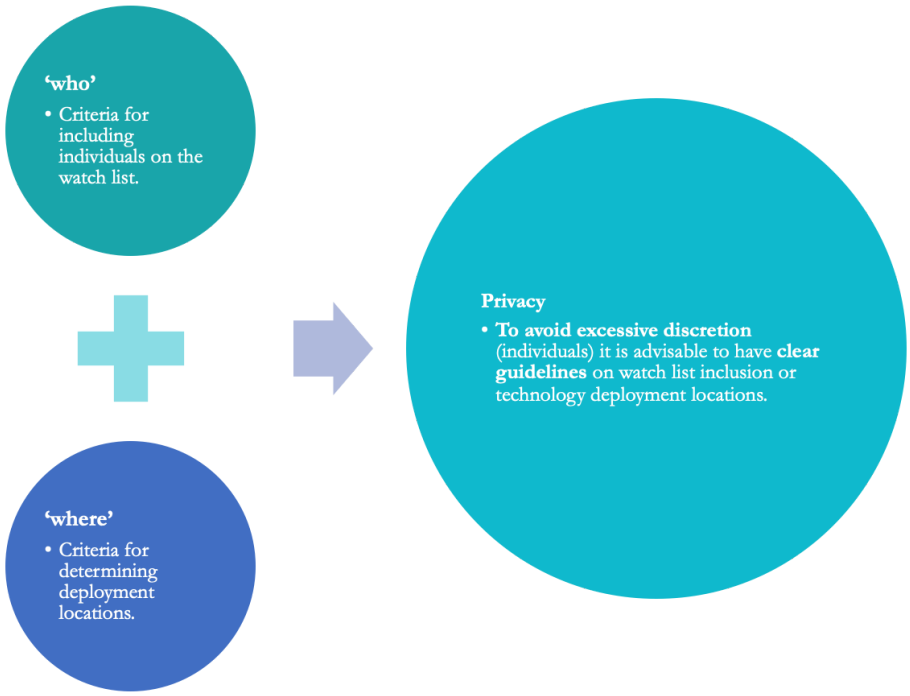


Figure 12. Main concerns regarding the implementation and use of AFR technology within organisational processes. By Lorena Flórez based on the R v SWP case.

[259] EWCA Civ 1058, Case R Bridges v CC South Wales 2020 [para 84].
[260] Ibid. [para 153].
[261] Ibid. [para 90-91].
[262] Ibid. [para 210].

France- CJEU: Ligue des droits humains ASBL v Conseil des ministres: Request for a preliminary ruling from the Cour constitutionnelle

In this case the CJEU addressed critical issues surrounding the use of automated systems for processing Passenger Name Record (PNR) data.[263] The CJEU found the PNR Directive is compatible with the ECHR, but required Member States to interpret key provisions in ways that differ from the Directive's original wording. The ruling highlights several important aspects of organisational embedding of such systems.

Firstly, the Court acknowledged the inherent margin of error in automated analyses of PNR data, stemming from the use of unverified airline data and predefined assessment criteria.[264] This recognition underscores the importance of robust organisational processes to mitigate these inaccuracies. Second, the Court emphasised that the system's effectiveness heavily relies on successful manual review and verification of positive results from the automated analysis.[265] This highlights a critical point of human-machine interaction within the organisational structure, where human oversight becomes crucial in ensuring the system's accuracy and fairness. In addition, the Court's insistence on a manual review process to identify and eliminate false positives and prevent discriminatory outcomes points to a necessary layer of human oversight within the organisational workflow.[266] Thus, this requirement implicitly acknowledges that the embedding of automated systems within organisations must include **clear processes for human intervention** and error correction.

In this regard, the Court stated that self-learning AI systems (ML) that can be modified without human intervention are deemed incompatible due to lack of certainty for human reviewers and data subjects. Because, opacity in AI systems prevents understanding the reasons for positive matches, potentially infringing on the right to an effective judicial remedy.[267] Thus, the Court establishes transparency rights for data subjects to support their right to an effective remedy in administrative procedures, as well as in judicial proceedings, where both the individual and the court must have the opportunity to examine all grounds, evidence, predetermined assessment criteria, and the operation of the programs applying these criteria.

Furthermore, the Court's ruling that enforcement agents must follow clear and precise rules set by Member State Passenger Information Units (PIUs) addresses the need for organisational guidelines and standardised procedures.[268] It stated that the directive for PIUs should also document the manual review process and all other PNR data processing activities pointing to the need for robust internal oversight mechanisms. This requirement for

[263] CJEU, Case C-817/19, *Ligue des droits humains ASBL v Conseil des ministres,* ECLI:EU:C:2022:491.
[264] Ibid. [para 106].
[265] Ibid. [para 124].
[266] Ibid. [para 203].
[267] Article 47 of the Charter. See also Palmiotto, F. (2024). When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis. *German Law Journal, 25(210).*
[268] Ibid. [para 205-206].

documentation creates an audit trail, enabling both internal and external oversight of the system's operation within the organisation.[269] While the Court's emphasis on establishing objective review criteria highlights the need for organisations to develop and implement clear standards for human oversight.[270] An advocacy group such as EDRi considers that the Court's approach grants too much trust to Member States and does not sufficiently address the high error rates in automated analysis.[271]

Therefore, while the Court's decision provides important guidelines for the embedding of automated PNR data processing systems within organisations, it also reveals potential gaps in oversight, which will be discussed in chapter 4 of this Handbook. Organisations implementing such systems will need to go beyond mere compliance with these guidelines and establish comprehensive, ongoing oversight mechanisms to ensure the fair and effective operation of these systems over time.

3.6 Liability and effective remedies

The continuous development of AI systems and algorithms and their advantages in ensuring quick and effective decision-making promote their usage in public administration. However, the use of these systems could lead on the one hand to an adverse impact on fundamental rights, and on the other hand to damage, for example, due to the risk of bias and loss of human control over the systems and in terms of the AI liability Directive 'damage caused by an output of an AI system or the failure of such as system to produce an output where such an output should have been produced'.[272] This section will focus on instances in which AI usage has led to damage or adverse impact on fundamental rights, the possibility of liability, and other effective remedies.

3.6.1 Use of AI and Adverse Impact on Fundamental Rights

The use of AI and algorithms in public administration often contributes significantly to adverse impacts on fundamental rights. Problems arise, amongst others, from the lack of algorithm transparency, lack of contestability, bias, and other reasons for the unfairness of AI systems, which may lead to discrimination or privacy and data protection issues.[273]

---

[269] Ibid. [para 207]. See also Naarttijärvi, M. (2023). AI and Sensitive Personal Data Under the Law Enforcement Directive: Between Operational Efficiency and Legal Necessity' in E Gill-Pedro and A Moberg (eds), *Yearbook of Socio-Economic Constitutions 2023*, *Springer, Cham.*

[270] Ibid. [para 206].

[271] European Digital Rights (EDRi). (2020). Mass Surveillance of External Travellers May Go on, Says EU's Highest Court.<https://edri.org/our-work/mass-surveillance-of-external-travellers-may-go-on-says-eus-highest-court/>.

[272] Dor, B and Coglianese, C. (2021). Procurement as AI Governance, *IEEE Transactions on Technology and Society, 2(4)*, 192-199.

[273] Rodrigues, R. (2020). Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities. *Journal of Responsible Technology, 4(100005)*. 5-7.

Considering the previous cases and application discussed above such as the the profiling system of unemployed people, and the Gotham system in relation to other potential applications of AI systems in the public sector —such as profiling for job markets, university admissions, social benefits allocation, or even for translating or filing public documentation— reveals the complex ethical and legal landscape surrounding these technologies.[274]

The use of AI systems in ways similar to those illustrated in the Gladsaxe and Gotham cases has been addressed in the EU AI Act, specifically in Article 5(1)(c). This article prohibits public authorities from using AI systems to evaluate or classify the trustworthiness of individuals when such systems result in harmful or unfair treatment of individuals or groups in social contexts unrelated to the original purpose for which the data was collected, or when the treatment is unjustified or disproportionate to the individual's social behaviour or its severity.[275]

This prohibition reflects growing concerns about the potential for AI systems to perpetuate or exacerbate existing social inequalities and biases, especially when deployed by public authorities in contexts that significantly impact citizens' lives and opportunities. For instance, an AI system used for job market profiling might inadvertently discriminate against certain groups based on historical data that reflects past biases in hiring practices. Similarly, an AI system for university admissions might disadvantage students from certain socioeconomic backgrounds if it relies too heavily on factors indirectly linked to wealth or social status.

Therefore, as we consider the implementation of AI systems in public administration, it is crucial to ensure that governance frameworks not only address the specific prohibitions of the regulatory framework but also uphold the fundamental right to an effective remedy. This may involve developing new procedures for algorithmic transparency, creating specialised bodies to review AI-driven decisions, and ensuring that legal and administrative systems are equipped to provide meaningful redress in cases involving AI.

3.6.2 The need for remedies

The right to an effective remedy is a fundamental principle of European law, enshrined in Article 47 of the Charter of Fundamental Rights. This article guarantees individuals the right to an effective remedy and a fair trial by outlining several essential criteria: the remedy must

[274] Rachovitsa, A. and Johann, N. (2022). The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case. *Human Rights Law Review, 22(2)*. 12-15.
[275] Appelman, N. and others. (2021). Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law, 12(257)*. 259-262.

be effective, the public hearing must occur within a reasonable time, an independent and impartial tribunal must be established by law, and there must be provisions for individuals to be advised, defended, and represented.[276] This right is crucial in ensuring that all rights established in the Charter are protected. Furthermore, it mandates that both national and EU courts must actively protect these rights in practice and in law.[277]

The integration of AI and algorithmic models into public administration, as discussed earlier, brings challenges that necessitate effective remedies—a guarantee provided by this article at the EU level. This section explores the possibilities for legal liability and other remedies to address the harms caused by AI in public administration.

3.6.3 Legal Liability and the Claim for Damages

Legal liability concerning AI can sometimes be ineffective or incomplete, leading to liability gaps that prevent injured parties from obtaining redress. These gaps often arise when traditional liability regimes, designed for human actors, are applied to the novel challenges posed by AI.[278] This section examines how these gaps manifest in three types of machine behaviour—individual, hybrid, and collective—and how alternative liability regimes could provide effective remedies and support claims for damages by AI system users. The previous analysis is based on the classification proposed by Anna Beckers and Gunther Teubner in their book on AI liability regimes. It is important to note that this is one perspective among many in the rapidly evolving field of AI governance and liability.[279] Their book develops this threefold typology based on information technology studies, distinguishing between **individual, hybrid, and collective machine behaviour.** The authors argue that these specific risks require tailored legal responses. They propose recognizing personified algorithms as vicarious agents, human-machine associations as collective enterprises, and interconnected systems as risk pools, with corresponding liability rules for each.

In **individual machine behaviour**, the risk inherent in the algorithmic system stems from its single source code or interactions with the environment. Here, the liability regime focuses

[276] Piątek, W. (2019). The Right to an Effective Remedy in European Law: Significance, Content and Interaction. *China-EU Law Journal, 6*. 165-167.
[277] Demková, S. (2023). Automated Decision-Making and Effective Remedies : The New Dynamics in the Protection of EU Fundamental Rights in the Area of Freedom, Security and Justice. *Edward Elgar Publishing Limited*. 25-30.
[278] De Conca, S. (2022). Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions: in Bart Custers and Eduard Fosch-Villaronga (eds), *Law and Artificial Intelligence*, *5*.
[279] Beckers, A. and Teubner, G. (2022). Three Liability Regimes for Artificial Intelligence Algorithmic Actants, Hybrids, Crowds.

on the legal personality of the algorithm.[280] In hybrid human-machine behaviour, where close interactions occur between humans and machines, the liability regime must account for this association by attributing a quasi-actor status to the human-machine collaboration. Collective behaviour, resulting from interconnected machine systems, necessitates a liability regime centred on the distribution of responsibility, leading to a 'de-individualized' approach to liability.

The autonomy risk associated with individual machine learning highlights the challenges of identifying responsible parties when damage arises from uncontrollable AI behaviour. This risk includes difficulties in pinpointing the decision-making agent, misunderstandings between human operators and AI, reduced productivity, and unpredictable algorithmic behaviour. The current liability regime, which relies on contractual liability, often fails to address these issues adequately.[281] A potential solution is to adopt vicarious performance, which closes the liability gap by holding the human principal accountable and limiting the legal personhood of the AI system.

Non-contractual liability regimes also face challenges in addressing damages caused by autonomous systems. However, adopting vicarious liability in tort, which limits the legal personhood of the system, can help close the liability gap by holding the user—who deploys and benefits from the AI—solely responsible.

Hybrid behaviour presents a unique risk of association, complicating the determination of individual responsibility.[282] Enterprise liability offers a viable solution by treating human-machine interactions as bilateral relationships or quasi-associations, allowing for distributed liability among all concerned parties, such as the user, producer, and programmer. Victims can claim damages from the actor with the most economic benefit, while the liable actor can seek redress from other responsible parties.[283]

The interconnectivity risk associated with **collective behaviour** makes it difficult to identify liable actors and foresee damage, as the outcomes are the result of 'systemwide' behaviour across multiple machine systems.[284] A 'no-fault fund-based system' is the most appropriate approach to address this risk. This solution aims to close the liability gap and ensure damage recovery by establishing a regulatory process that interacts with existing insurance and

[280] Ibid. 23-27.
[281] Ibid. 139-141.
[282] Ibid 89-110.
[283] Ibid. 153-158.
[284] Ibid. 111-135. See also Erdelyi, O. and Erdelyi, G. (2021). The AI Liability Puzzle and A Fund-Based Work-Around. *Journal of Artificial Intelligence Research, 70(1309),* 1312.

liability laws. It involves determining ex-ante the potentially liable actors based on their economic benefit and ability to pay damages, and ex-post, distributing the fund to victims.

Therefore, the classification proposed by Beckers and Teubner offers a comprehensive framework for addressing AI liability through three regimes: vicarious liability for autonomous software agents, enterprise liability for human-AI interactions, and collective fund liability for interconnected AI systems. This approach aims to address the specific risks associated with different types of machine behaviour: digital autonomy, double contingency, and opaque interconnections. By recognizing personified algorithms as vicarious agents, human-machine associations as collective enterprises, and interconnected systems as risk pools, the framework provides a nuanced approach to AI governance and liability.

In conclusion, when applying this framework to AI implementation in public administration, it becomes evident that transparency and governance are crucial. Public bodies must establish clear policies for disclosing information about AI systems, including decision automation, associated risks, and technical management aspects. This aligns with the authors' emphasis on recognizing the **specific risks and socio-technical configurations of AI systems.** By adopting such an approach, public administration can not only ensure compliance with legal standards but also enhance public trust. This transparency and accountability in AI-driven decision-making processes address the concerns raised by Beckers and Teubner about the responsibility gap in AI systems, while also meeting the practical needs of public administration. Ultimately, this combined perspective offers a pathway to responsible AI implementation that balances legal, ethical, and practical considerations in the public sector.

3.6.4 Injunctions and Invalid Administrative Acts

The implementation of AI in public administration presents unique challenges for judicial oversight and remediation. This section explores two key judicial mechanisms—injunctions and administrative act review—and their application in the context of algorithmic governance.

Injunctions serve as a crucial tool in the judicial arsenal for addressing potential harms arising from the use of algorithms or automation in public administration.[285] These court orders, which can be permanent, temporary, or preliminary, are particularly relevant in the early stages of AI implementation when the full impacts of algorithmic decision-making may not be fully understood.[286] As an example, in the case of *Ark. Dep't of Human Servs. v.*

[285] Injunction (LII / Legal Information Institute) <https://www.law.cornell.edu/wex/injunction>.
[286] Sir Richard, A. (2022). Injunctions in European Law in: Judicial Reflections. 65-69.

Ledgerwood, the court took a decision to issue an injunction until the Department of Human Services could explain the rationale behind its automated decision.[287] Even in different legal frameworks, this case illustrates the potential for injunctions to act as a safeguard against opaque or potentially harmful AI implementations in public administration, highlighting the judiciary's role in ensuring **transparency and accountability** in algorithmic governance.[288] In addition, the Ledgerwood case also illustrates the potential for AI systems to significantly impact vulnerable populations, as evidenced by the Resource Utilisation Group (RUG) system's role in determining care allocations for disabled individuals.[289] This raises important questions about the **equity and fairness** of AI implementations in public services.[290]

Another example is the legal battlebetween *noyb* against the credit reference agency CRIF GmbH and the address trader AZ Direct.[291] Noyb claims that AZ Direct unlawfully provides personal data (e.g., name, address, date of birth, gender) to CRIF, which uses this data to calculate individuals' creditworthiness, a purpose not legally permitted. The Austrian Data Protection Authority (DSB) has confirmed that this practice violates GDPR principles, specifically the principle of purpose limitation.[292]

According to *noyb*, despite DSB's findings, no measures have been taken to stop the illegal data processing. As a result, *noyb* has filed a lawsuit at the Vienna Regional Court for Civil Matters on behalf of seven affected individuals. The lawsuit aims to secure an injunction to stop CRIF and AZ Direct from continuing their illegal data processing practices. The lawsuit also seeks non-material damages for the affected individuals and restitution of profits unlawfully earned by the companies through these practices. This case underscores the importance of **explainability in AI systems** used in public administration, a principle that aligns with the broader discourse on ethical AI and the right to explanation enshrined in regulations such as the GDPR.

---

[287] Ark. Dep't of Human Servs. v. Ledgerwood, 571 S.W.3d 911, 2019 Ark. 121. While this case falls under U.S. jurisdiction, it provides valuable insights into how injunctions can serve as an effective remedy in the context of automated decision-making systems.
[288] Calo, R. and Keats Citron, D. (2020). The Automated Administrative State: A Crisis of Legitimacy. *Emory Law Journal, 70,* 797-799.
[289] Ibid. 780.
[290] This point is explained and mentioned by scholars such as Virginia Eubanks, (2018). Automating Inequality. *St Martin's Press, Inc Subs of Macmillan Publishers.* <https://dl.acm.org/doi/book/10.5555/3208509>.
[291] NOYB – European Center for Digital Rights vs CRIF GmbH and the address trader AZ. Direct.<https://noyb.eu/sites/default/files/2023-11/Schadenersatz-%20und%20Unterlassungsklage_geschw%C3%A4rzt.pdf>.
[292] DSB (Austria), 2023-0.592.319. ECLI:AT:DSB:2024:2023.0.592.319 [par. 11] <https://gdprhub.eu/index.php?title=DSB_(Austria)_-_2023-0.592.319>.

The **judicial review of administrative acts** provides another avenue for ensuring accountability in the use of AI by public administrations. This mechanism allows courts to declare administrative acts invalid if they are found to be arbitrary, capricious, or not in accordance with the law. In the context of AI implementation, this review process takes on added complexity due to the often opaque nature of algorithmic decision-making.[293]

For instance, the Dutch administrative court's approach, which considers principles of due care and adequate reasoning, offers a model for how judicial review might be adapted to address AI-driven administrative acts.[294] However, as noted, the effectiveness of such reviews hinges on the judiciary's ability to critically analyse automated decision-making processes. This highlights the need for enhanced technological literacy among judges and legal professionals.[295] In this regard, the absence of clear mandates and standards for AI implementation in public administration poses a significant challenge to effective judicial review.[296] This gap underscores the need for interdisciplinary collaboration between legal experts, technologists, and policymakers to develop comprehensive frameworks for AI governance in the public sector.

The judicial mechanisms discussed here have significant implications for the implementation of AI in public administration. First, in regards to the principles of transparency and explainability, the potential for injunctions and judicial review underscores the importance of developing AI systems that are transparent and explainable, enabling both administrators and the judiciary to understand and scrutinise automated decision-making processes. Second, public administrators must implement robust procedural safeguards to ensure that AI-driven decisions can withstand judicial scrutiny, including mechanisms for human oversight and intervention. Third, given the evolving nature of AI technologies, public administrations must establish systems for continuous monitoring and evaluation of AI implementations to preempt potential legal challenges. Fourth, there is a pressing need for capacity building within both public administration and the judiciary to enhance understanding of AI technologies and their implications for administrative law. Finally, the challenges highlighted by these judicial mechanisms point to the need for comprehensive regulatory frameworks specifically tailored to AI implementation in public administration.

---

[293] Unzen, G. (2023). Artificial Intelligence and the Administrative State: Regulating the Government Use of Decision-Making Technology Note. *Minnesota Journal of Law, Science and Technology, 25(209).*

[294] Barkhuysen, T. and others. (2012). The Law on Administrative Procedures in the Netherlands. *Netherlands Administrative Law Library.* <https://www.nall.nl/tijdschrift/nall/2012/06/NALL-D-12-00004>.

[295] This point is emphasised by scholars such as Pasquale, Frank. (2016). Black Box Society: The Secret Algorithms That Control Money and Information.

[296] Unzen, G. (2023). Artificial Intelligence and the Administrative State: Regulating the Government Use of Decision-Making Technology Note. *Minnesota Journal of Law, Science and Technology, 25(209).*

In conclusion, while injunctions and administrative act review provide important judicial tools for addressing the challenges of AI implementation in public administration, their effectiveness is contingent upon broader systemic changes. These include enhancing technological literacy within the legal system, developing clear standards for AI governance, and fostering a culture of transparency and accountability in algorithmic decision-making. As public administrations continue to adopt AI technologies, the evolution of these judicial mechanisms will play a crucial role in shaping the landscape of algorithmic governance and ensuring that the benefits of AI are realised without compromising fundamental rights and democratic principles.

3.7 Concluding remarks: Organisational embedding stage

Embedding AI in public administration is a complex process that requires careful consideration of both technical and ethical aspects. Public bodies must ensure that AI systems integrate seamlessly with existing technologies, scale effectively with growing demands, and align with the specific needs of the organisation. However, technical success alone is insufficient; the broader implications for fundamental rights, legal compliance, and ethical governance must also be carefully managed. For AI to serve its intended purpose in public administration, these elements must be addressed simultaneously.

A critical aspect of embedding AI systems is the need for **transparency and accountability**. Public entities must provide clear, understandable explanations to citizens regarding how AI-driven decisions are made, including how different factors are weighted in these decisions. Additionally, it is essential to establish clear governance frameworks that prevent discriminatory outcomes, ensure compliance with data protection laws, and maintain the integrity of the decision-making process. These frameworks should include mechanisms for citizens to challenge decisions and seek redress, thus fostering public trust and ensuring that AI systems operate within the law.

Finally, the embedding process should prioritise **continuous monitoring and improvement** of AI systems. This involves not only addressing technical issues as they arise but also adapting to evolving societal needs and technological advancements. By implementing robust feedback loops and mechanisms for human oversight, public administrations can identify and mitigate potential risks to fundamental rights, such as discrimination or privacy violations. This ongoing process is essential to maintaining public confidence in AI-powered systems, ensuring they remain aligned with democratic values and serve the best interests of all citizens.

3.7.1. Guiding questions

The aim of this handbook is to equip participants with practical information and real-world case studies to help them mitigate potential conflicts with fundamental rights during AI system development. In this section, we highlight some of the most critical questions that public entities should consider during the organisational embedding stage, ensuring that their AI initiatives are both effective and aligned with legal and ethical standards.

Fundamental Rights

- Which fundamental rights, such as non-discrimination, equal treatment, privacy, data protection, and human dignity, could be impacted by the AI system?
- How might the system inadvertently perpetuate or amplify existing biases, and what safeguards are in place to prevent this?
- Does the organisation have mechanisms for ongoing monitoring and review of the algorithm's outcomes to ensure alignment with fundamental rights?
- What procedures are in place to allow citizens to challenge AI-driven decisions and seek redress?

Organization's Values and Culture

- Is the AI system interoperable with existing organisational infrastructure, scalable to meet future demands, and aligned with the organisation's core values?
- In what ways might the organisational culture influence the AI system's recommendations and outcomes?
- How is personnel training structured to address and mitigate ethical risks and systemic biases associated with AI use?
- What protocols and procedures are established to guide manual reviews of AI decisions, ensuring human oversight?
- Are diverse stakeholders actively involved in the ongoing review, monitoring, and governance of the AI system to ensure comprehensive oversight?

4. Public decision-making using AI systems

Lorena Flórez, Ida Varošanec, Lucas Haitsma, Linda Ardenghi, Ivo Emanuilov, Ginka Stoyanova, Alexandra Tsvetkova, Tobias Nowak, Alma Onaie, Larisa-Diana Bularca

4.1 Risks for fundamental rights in the decision-making using AI systems

The integration of AI and automated decision-making systems in public administration has ushered in a new era of governance, promising enhanced efficiency and effectiveness across various sectors. From international to national levels, governments are increasingly adopting these technologies to address complex challenges in areas such as welfare fraud detection, resource allocation, and law enforcement. For instance, AI-driven approaches to resource allocation have revolutionised public service provision, optimising the distribution of resources in key sectors such as healthcare, education, and transportation. These innovations promise to increase efficiency and effectiveness in public service delivery. However, as these systems become more deeply embedded in the fabric of public administration, it is crucial to examine their long-term implications, particularly in the context of public decision-making

The third stage, the public decision-making stage, refers to the long-term use of an algorithmic or AI system in practice after its design and embedding within a governmental organisation.[297] This stage is characterised by the ongoing use of AI tools, where decisions affecting citizens' lives are made with different degrees of algorithmic influence. The potential risks that may have gone unnoticed during earlier stages can manifest here, potentially leading to opaque and discriminatory outcomes if not properly addressed.[298] For example, using a complex discriminatory algorithmic system coupled with a lack of adequate human intervention and oversight mechanisms can lead to opaque discriminatory decisions being taken.[299] As such, it is imperative that mechanisms for explanation, justification, and transparency are firmly in place, allowing citizens to understand the decision-making process and invoke their rights when necessary.

Crucial to this stage is, firstly, the explanation, justification, and transparency of the decision-making process and the role of the algorithmic system within this process as this

[297] Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, *30*(3), 101683, pp. 6-7.; See also Haitsma, L. (2023). The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy. *Recht der Werkelijkheid*, *44*(2), 61-83.

[298] The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy. *Recht der Werkelijkheid*, *44*(2), 61-83.

[299] McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. *International & Comparative Law Quarterly*, *68*(2), pp. 323-324.; Lazcoz, G., & De Hert, P. (2023). Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities. *Computer Law & Security Review*, *50*, 105833.

enables citizens to invoke their fundamental rights.[300] Second, concepts of ongoing oversight, evaluation, maintenance, and support of the AI system (*revision*). This includes monitoring the system's performance, addressing any issues or errors that may arise, and continuously updating the system to adapt to changing needs and emerging technologies. Effective maintenance and support are crucial to ensuring the long-term viability and reliability of the AI-powered tools within the organisation. Additionally, these mechanisms ensure a cycle of feedback and adjustments in the design, implementation, and use of the system.[301] Such mechanisms contribute to fundamental rights issues, such as discrimination, being identified and remedied.

This section of the handbook provides a comprehensive examination of the decision-making stage in the AI life cycle, with a particular focus on its implications for fundamental rights. We will begin by explaining the decision-making stage, detailing its components and the technological issues involved. This will be followed by an in-depth exploration of fundamental rights issues that arise during implementation, illustrated through real-world examples and case studies from various jurisdictions. Additionally, we will outline the legal, regulatory, and ethical frameworks governing AI decision-making in public administration. This analysis will be structured across multiple levels, with particular attention given to ethical principles such as accountability, redress, and revision, and how these principles are reflected in soft law and policy frameworks across different governance levels. Furthermore, this section will address the critical issues of liability and effective remedy in the context of AI decision-making, providing insights into how legal systems are adapting to these new challenges. Finally, we will offer practical guidelines for judges to assist them in safeguarding fundamental rights during the decision-making stage of AI systems in public administration.

4.2 Decision-making stage of AI life cycle

The decision-making stage of the AI life cycle is a critical phase that involves continuous operation of AI tools within an organisation's decision-making processes (*monitor and supervision*). This stage encloses several key technological, and governance considerations.

From a ***technological perspective***, this stage involves evaluation of AI models to assess their performance and fairness in real-world settings (*through internal or external audits*). An adequate evaluation process should encompass not only technical metrics like operating characteristics,[302] receiver operating characteristic (ROC) curve, and area under the curve

[300] Ződi, Z. (2022). Algorithmic explainability and legal reasoning. *The Theory and Practice of Legislation*, *10*(1), p. 69, 80, 89-91.

[301] Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, *30*(3), 101683. Pp. 6-7 & Yeung, D., Khan, I., Kalra, N., & Osoba, O. (2021). *Identifying systemic bias in the acquisition of machine learning decision aids for law enforcement applications*. RAND.p. 20; Tilburg University, Non-Discrimination by Design. Tilburg University 2021, pp. 61-68; Haitsma, L. (2023). The Murky Waters of Algorithmic Profiling: Examining discrimination in the digitalized enforcement of social security policy. *Recht der Werkelijkheid*, *44*(2), 61-83.

[302] Performance metrics are a part of every ML pipeline. These metrics show the progress of the model, and put a number on it. All ML models, whether it is linear regression, or a SOTA technique like BERT, need a

(AUC),[303] but also address broader implications of fairness and integration within the organisation's workflows.[304] For instance, ensuring model calibration and assessing performance across various populations (*testing before deployment*) are crucial steps.[305]

From a ***governance perspective***, this stage addresses, first, transparency obligations. Deployment requires a deep understanding of how these models perform in real-world settings, how public servants interpret their outputs, and their overall public utility.[306] Transparency, generalizability, and fairness are key concerns during deployment, as any deficiencies in these areas can have significant real-world consequences. For example, a public health agency using an AI system to allocate resources during a pandemic should publish reports explaining how the AI determines resource distribution, including the criteria used and how these decisions align with public health goals. This transparency (*explainability*) helps build trust and ensures that the public and stakeholders understand the decision-making process.[307]

Second, it is essential to document all stages of the model development and make available the evaluation, audit, and internal or external monitoring processes of the model.[308] For

---

metric to judge performance. For instance, to evaluate the performance of a classification model, different metrics are used such as: accuracy, confusion matrix, precision, recall and others. See for instance Thomas, R. and Uminsky, D. (2020). The Problem with Metrics Is a Fundamental Problem for AI. *EDSC (Ethics of Data Science Conference)*. <https://arxiv.org/abs/2002.08512v1>.

[303] The area under the ROC curve (AUC) represents the probability that the model, if given a randomly chosen positive and negative example, will rank the positive higher than the negative. See for instance Reyna, M. and Clifford, G. (2022). Rethinking Algorithm Performance Metrics for Artificial Intelligence in Diagnostic Medicine. *JAMA 329*.

[304] Cao, Z. and others. (2021). Performance Metrics for Artificial Intelligence (AI) Algorithms Adopted in Prognostics and Health Management (PHM) of Mechanical Systems. *Journal of Physics: Conference Series (IOP Publishing 2021)*. <https://iopscience.iop.org/article/10.1088/1742-6596/1828/1/012005>.

[305] De Silva, D. and Alahakoon, D. (2022). An Artificial Intelligence Life Cycle: From Conception to Production. *Patterns, 3*. 6-7.

[306] Madelena, Y. and others. (2022). The AI Life Cycle: A Holistic Approach to Creating Ethical AI for Health Decisions. *Nature medicine, 28*. 2247-2248.

[307] As reference, the implementation of Brazil's Emergency Aid (Auxilio de Emergencia) program, designed to mitigate the socio-economic impacts of the COVID-19 pandemic, highlights significant gaps in transparency and rights protection, particularly for the most vulnerable populations. While the program relied heavily on automated data processing by Dataprev to determine eligibility, it failed to provide citizens with crucial information about the decision-making criteria, the rationale behind the use of such a system, or mechanisms to appeal decisions. This lack of transparency and access to information left beneficiaries without the means to understand or challenge the outcomes of the automated processes, raising concerns about the protection of their rights in a context of pervasive inequality. For more information about this case see: Tavares, C. and others. (2022). BRASIL | Auxilio de Emergencia – Inteligencia Artificial e Inclusión En América Latina. *Derechos Digitales & InternetLab*. <https://ia.derechosdigitales.org/casos/brasil-auxilio-de-emergencia/>; See also Stephan Grimmelikhuijsen, 'Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making' (2023) 83(2) Public Administration Review 241-262.

[308] Laato, S. and others. (2022). AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering. *Proceedings - 1st International Conference on AI Engineering - Software*

instance, a city's transportation department deploying an AI system for traffic management should maintain detailed records of the system's development, including data sources, model training, and testing processes. These records should be accessible for external audits and public scrutiny to ensure that the system operates fairly and effectively (*accessibility*).[309]

Finally, as an integral part of the decision-making stage, it is crucial that the organisation ensures that the end users of the AI systems have access to both *ex ante* and *ex post* mechanisms to safeguard rights such as the right to explanation, the right to appeal and redress, and the right not to be subject to solely automated decision-making.[310] For example, a social services agency using AI to determine eligibility for welfare benefits should have a clear process in place allowing individuals to challenge decisions, request explanations for how decisions were made, and appeal if they believe the AI has made an error (*accountability*). This ensures that individuals' rights are protected and that the AI system operates within ethical and legal boundaries.

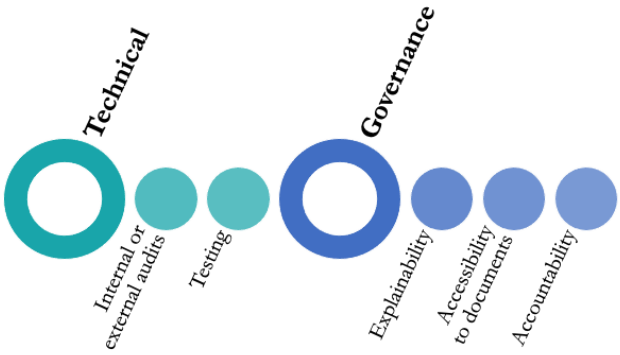


Figure 12. Elements of the decision-making stage of AI systems. By Lorena Flórez.

Engineering for AI, CAIN. 113; McGregor, L., Murray, D. and Ng, V. (2019). International Human Rights Law as a Framework for Algorithmic Accountability. *International & Comparative Law Quarterly, 68(309)*, 311.

[309] One example of accessibility is the implementation of algorithm registers as the most basic form of informing the people about the list of the algorithms used by governments. Algorithm registers have been developed at national, subnational, and municipal levels of government. For more information about these registers see: van Vliet, M. and others. (2024). Defining and Implementing Algorithm Registers: An Organizational Perspective. *ECIS 2024 Proceedings* <https://aisel.aisnet.org/ecis2024/track04_impactai/track04_impactai/5>; see also Sengupta,U. (2022). Algorithm Registers as an Infraestructural Element for Transparent Government' in A Brandusescu and J Reia (eds), *Artificial intelligence in the city: Building civic engagement and public trust.* Centre for Interdisciplinary Research on Montreal, McGill University; Schuitemaker, N. and others. (2024). Realizing the Accountability of Algorithms in the Public Sector: A Reference Method for Managing Algorithm Registers. *511 LNBIP Lecture Notes in Business Information Processing 325*.

[310] Novelli, C., Taddeo, M. and Floridi, L. (2023). Accountability in Artificial Intelligence: What It Is and How It Works. *AI and Society 1*.

4.3 Fundamental rights issues in the decision-making stage- Study cases of real applications

In the following section, we explore the decision-making processes of AI systems through the lens of real-world case studies and applications. We consider key elements such as internal and external audits, adequate documentation of processes, accessibility of these documents by diverse stakeholders, and accountability measures. Our analysis focuses on how these issues have manifested in the practical deployment of AI tools in the public sector. These case studies highlight the diverse ways in which decisions adopted by organisations have impacted citizens and their fundamental rights. We examine how these decision-making processes may have violated or upheld fundamental rights, providing concrete examples of the intersection between AI implementation and civil liberties.

Furthermore, where real-world examples may be limited or where emerging technologies are involved, we also analyse potential implications for fundamental rights in hypothetical scenarios. This approach allows us to anticipate potential challenges and consider preemptive measures, even in situations that have not yet fully materialised in practice.

4.3.1. EU case law

Germany - CJEU: SCHUFA Holdings

German credit reference agency, SCHUFA, used automated processes to generate credit scores, which were then used by lenders to make decisions about credit applications (*risk assessment*). The case concerned the refusal of SCHUFA, to disclose the various elements used to assess the creditworthiness of a client, upon their request.[311] After considering the scope of Article 22(1) GDPR and the facts of the case, the Court established that Article 22(1) GDPR was indeed applicable to the case.[312] The Court additionally noted that, in almost all cases, when an insufficient probability value, meaning a low credit score, was assigned to a consumer applying for a loan to a bank, the loan was refused.[313]

Although the SCHUFA case **is centred on a private company**, it provides valuable insights for the implementation of AI systems within public administration. When public bodies use AI systems, particularly those relying on third-party assessments or data, it is crucial to establish robust governance structures. These structures should clearly define how much influence third-party reports or assessments have on decision-making processes and ensure that these systems do not operate in a black box.[314] Transparent processes must be in place,

[311] CJEU, Case C-634/21, *SCHUFA Holding AG v. Hesse* ECLI:EU:C:2023:987. [14-16]
[312] Ibid. [47-49].
[313] Ibid. [48].
[314] Ibid. [54 and 56].

allowing citizens to understand **how decisions are made**, **seek redress if necessary, and ensure that their rights are protected**.[315]

In response if a request for a preliminary ruling submitted by the Wiesbaden Administrative Court (Germany), the CJEU had to answer the question '*whether the establishment of a probability value by a credit information agency based on personal data relating to a person and concerning his or her ability to meet payment commitments in the future constitutes automated individual decision-making with the meaning of Article 22(1) GDPR?*'. To answer the question, the Court first analysed the applicability of Article 22(1) GDPR, which was based on 3 cumulative conditions: a decision has been made; that decision is based solely on automated processing including profiling; and it must produce legal effects concerning the interested party or similarly significant affect him or her.

Firstly, the Court considered the concept of 'decision' within the meaning of Article 22(1) GDPR, asserting that it was apparent that the concept referred not only to acts which produce legal effects concerning the person at issue, but also acts which similarly affect them, which is further confirmed in Recital 71. Thus, the Court established that the concept of 'decision' was broad enough to encompass the result of calculating a person's credit score in the form of a probability value representing the ability of that person to respect their payments in the future. In fact, if a restrictive interpretation of Article 22 GDPR was adopted, the probability value would only be considered as a preparatory act and only the act adopted by the third party would be considered a decision within the meaning of Article 22 GDPR. This restrictive interpretation would then constitute the risk of creating a lacunae in the legal protection set by the Article. Secondly, it considered that activity such as that of SCHUFA, establishing a probability value to represent an individual's credit score, met the definition of profiling established in Article 4(4) GDPR.

Lastly, the Court considered the factual findings of the case and established that, in almost all cases, when an insufficient probability value, meaning a low credit score, was assigned to a consumer applying for a loan to a bank, the loan was refused. Consequently, it could be established that the assigning of a probability value affected the data subject significantly and that all three cumulative criteria of Article 22 GDPR were met. The establishment of a probability value by a credit information agency and referred to a bank played a determining role in granting a loan; as such, it must be considered as a **decision producing legal effects or similarly significantly affecting the data subject** for the meaning of Article 22(1) GDPR.

In the case of automated decision-making, the controller is also subject to additional information obligation under Article 13(2)(f) and Article 14(2)(g) GDPR. Moreover, under Article 15(1)(h), the data subject enjoys the right to **obtain meaningful information** from the controller about the logic involved, the significance and the expected consequences of such processing.

[315] Ibid. [5 and 66].

Therefore, just as the Court emphasised in the SCHUFA case, public administrations must implement safeguards, including the possibility for **human oversight, the right to challenge automated decisions**, and the obligation to provide clear, understandable information about how AI systems function. Article 22 GDPR, paragraphs 2(b) and 3, provide that safeguards must be taken to ensure the data subject's rights and freedoms, and ensure fair and transparent data processing.[316] These measures are essential to ensure that AI systems are used responsibly and fairly in public administration.

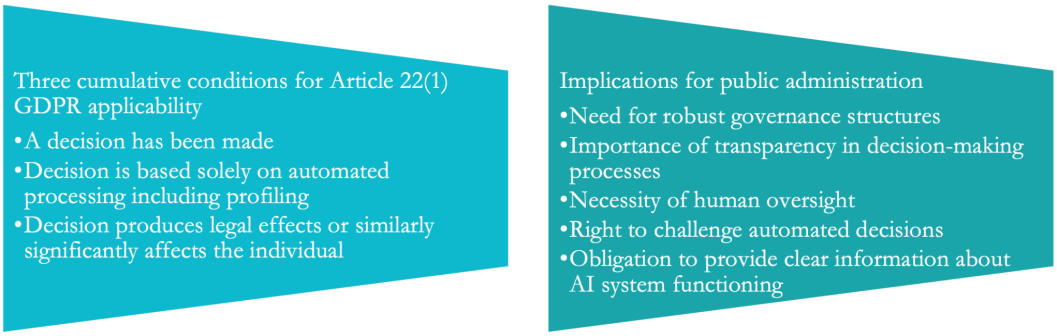


Figure 13. Applicability of Article 22(1) GDPR- to automated credit scoring. By Lorena Flórez based on the Schufa case.

France-CJEU: Ligue des droits humains ASBL v Conseil des ministres: Request for a preliminary ruling from the Cour constitutionnelle

In this case the CJEU addressed critical issues surrounding the use of automated systems for processing Passenger Name Record (PNR) data.[317] The CJEU found the PNR Directive is compatible with the ECHR, but required Member States to interpret key provisions in ways that differ from the Directive's original wording. The ruling highlights several important aspects from a decision-making stage of such systems.

One of the aspects of this case emphasises the critical role of **human oversight in the decision-making process** where automated systems, such as those processing PNR data, are involved.[318] The Court highlighted that decisions with legal consequences must prioritise manual review outcomes over automated processing, ensuring that **human intervention** acts as a safeguard against potential errors and biases in the system. This requirement for manual review is particularly significant given the inherent margin of error in automated analyses,

[316] Ibid. [54].
[317] CJEU, Case C-817/19, *Ligue des droits humains ASBL v Conseil des ministres,* ECLI:EU:C:2022:491.
[318] Ibid. [para 106]. See also Haitsma, L. M. (2023). Regulating algorithmic discrimination through adjudication: the Court of Justice of the European Union on discrimination in algorithmic profiling based on PNR data. *Frontiers in Political Science*, *5*, 1232601.

which often rely on unverified data and predefined assessment criteria.[319] By mandating that authorities must conduct these reviews without discrimination and provide affected individuals with **transparency regarding the assessment criteria** and systems used, the Court underscores the importance of procedural fairness and the right to an effective remedy.[320]

Moreover, the Court's decision outlines specific responsibilities for both the Passenger Information Units (PIUs) and national supervisory authorities in overseeing the processing of PNR data.[321] These bodies are tasked with ensuring that the use of automated systems does not lead to discriminatory outcomes, and that all processing activities are **documented and subjected to rigorous internal oversight**. The Court also explicitly ruled out the use of self-learning AI systems that operate without human intervention, citing the need for certainty and transparency for both human reviewers and data subjects. This ruling highlights the importance of having clear organisational guidelines and standardised procedures for the manual review process, which must be documented to create an audit trail for both internal and external oversight.[322]

However, while the Court's decision provides a robust framework for integrating human oversight into automated decision-making processes, it also reveals potential gaps in the oversight mechanisms. The decision does not explicitly require ongoing monitoring of system performance or regular audits of outcomes, which could lead to complacency in organisations that comply only with the minimum legal requirements. Additionally, the absence of directives to regularly review and update the predefined assessment criteria or algorithms could result in outdated or biassed systems over time. Therefore, to truly ensure fairness and effectiveness, organisations must go beyond mere compliance and establish comprehensive, continuous oversight mechanisms that adapt to changing circumstances and emerging challenges.

4.3.2. National courts

When an AI-assisted decision is rendered to an individual, they might need access to the feature values used to make the prediction, the weights of the specific elements in their case, or even more general weights in order to assert their rights. However, considerations of confidentiality can stand in the way of transparency, thereby making their reconciliation difficult.[323] While the CJEU has not yet provided specific guidance, national courts have been attempting to balance competing interests in different cases.

---

[319] Article 47 of the Charter. See also Palmiotto, F. (2024). When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis. *German Law Journal, 25(210).*

[320] CJEU, Case C-817/19. [para 124].

[321] Ibid. [para 205-206].

[322] Ibid. [para 206].

[323] Ida Varosanec. 'Silence is golden, or is it? Trade secrets *versus* transparency in AI systems' *(The Digital Constitutionalist*, 17 November 2022). Available at https://digi-con.org/silence-is-golden-or-is-it/.

Slovakia: eKasa case- Tax authority

The Slovak tax authority employed three distinct algorithms: a chatbot 'TAXANA', an enterprise resource planning (ERP) system 'eKasa', and a risk-scoring machine-learning algorithm 'AIS-R'. The controversy centred around the 'eKasa' system and its associated 'AIS-R' algorithm, which processes data to predict VAT fraud risks. The Slovak Constitutional Court ruled that transparency obligations must take precedence over trade secrets in AI-assisted decision-making.[324] The court also emphasised the importance of individual access to information about automated assessments for the right to effective remedy.[325]

A key issue was the lack of legal basis for the ML algorithm, which was implemented based on a governmental decision rather than established law. This situation is not unique to Slovakia, as many EU Member States lack specific legislative norms regulating such algorithms or have inadequate safeguards to protect taxpayers' rights.[326] The Slovak Supreme Court ruled in favour of the claimants, declaring that the eKasa system and its buyer identification feature **were incompatible with the principle of legality and the rights to privacy and data protection.**

First, the Court emphasised that automated assessment of individuals based on personal data, such as risk-scoring algorithms, cannot be implemented solely on administrative decisions due to their potential adverse impacts on taxpayers. It argued that transparency, protection of individuals, and collective supervision are essential for ensuring accountability in AI-driven public administration decisions.[327] Thus, the court highlighted that if "*the legislator wants to enable such automated assessment on the basis of personal data, he must ensure the fulfilment of constitutional guarantees, which the Constitutional Court has already detailed.*"[328]

Second, the case also raised concerns about the extensive data collection required by the eKasa system, particularly the requirement for unique identifiers for both sellers and buyers, potentially allowing the tax administration access to complete purchase histories of individuals. It criticised the GDPR framework for relying solely on individual rights and called for a more systematic approach to address public administration issues.[329] In addition, the Court expanded the definition of automated decision-making to include ***decision-assistance scenarios*** where ML outputs significantly influence subsequent human decisions.[330] This interpretation addresses a limitation in Article 22 of the GDPR, which

[324] Constitutional Court of the Slovak Republic PL., Judgement 492/2021 Z.z. of the ÚS 25/2019-117 (2021) (hereinafter 'eKasa'). <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2021/492/20211217>..
[325] See eKasa judgement, para 133.
[326] See eKasa judgement, para 26.
[327] See eKasa judgement, para 127.
[328] See eKasa judgement, para 142.
[329] See eKasa judgment, para 136; Martin Husovec, 'The Slovak Constitutional Court on Risk Profiling and Automated Decision-Making by the Tax Authority' (*husovec.eu*, 17 December 2021) <https://husovec.eu/2021/12/the-slovak-constitutional-court-on-risk-profiling-and-automated-decision-making-by-the-tax-authority/> accessed 9 September 2024.
[330] See eKasa judgement, para 128.

narrowly defines automated decision-making as decisions based solely on automated processing.

Third, regarding trade secrets, the court stated that the State must ensure transparency regardless of the source of the software and that this should be considered during the procurement process.[331] The judgement emphasises the need for explainability of AI systems and prioritises due process and access to information over intellectual property protection.

Lastly, the Court provided guidance on necessary safeguards for algorithmic governance in public administration. It emphasised the importance of the principle of legality, transparency, and informational self-determination. The Court stressed that the legislature must ensure the legitimacy and proportionality of technological solutions proposed for tax administration. The Court also highlighted the need for collective controls, both ex-ante and ex-post, to mitigate risks to taxpayers' fundamental rights.

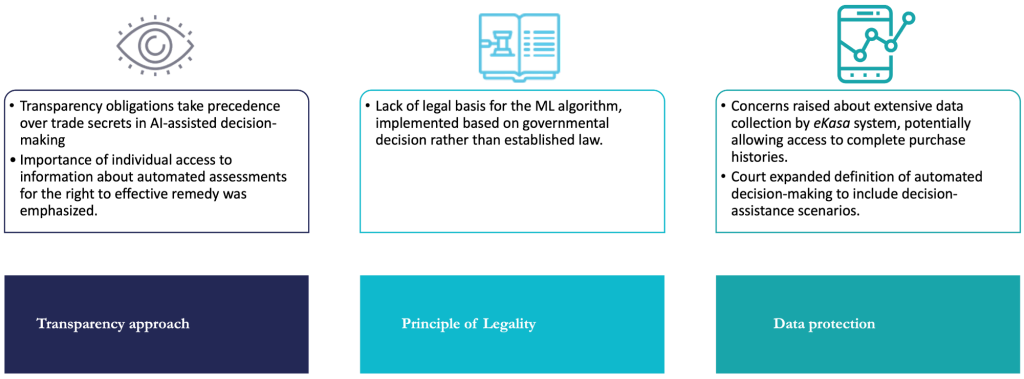


Figure 14. Analysis of the Slovak Constitutional Court: Controversy focused on eKasa and AIS-R, which predicted VAT fraud risks. By Lorena Flórez based on the eKasa case.

Italy

Recent developments in Italian administrative case law have demonstrated a significant shift in the judiciary's approach to the use of technology in administrative processes, particularly in relation to algorithms. Traditionally, Italian administrative courts, known as Tribunali Amministrativi Regionali (TAR), were perceived as conservative, especially in their hesitance to integrate technological innovations into legal interpretations.[332] However, recent rulings

[331] See also for more extensive analysis of the eKasa judgement: Hadwick, D. (2022). Behind the One-Way Mirror: Reviewing the Legality of EU Tax Algorithmic Governance. *EC Tax Review, 31(4)* <https://kluwerlawonline.com/journalarticle/EC+Tax+Review/31.3/ECTA2022019>. .

[332] For instance, in the decision no. 6606/2019, the administrative court of Rome declared that algorithm-based administrative decisions are **inherently illegitimate**. The case involved a mobility procedure for certain civil servants, such as school managers, which had been entirely handled by an algorithm. The court ruled that such procedures cannot be delegated to software alone, as algorithms inherently lack the capacity to uphold the principles of impartiality, effectiveness, and transparency that are essential to public administration. Administrative court of Rome, Section III bis, Decision no. 6606 of 27 May 2019.

indicate a departure from this perception, with courts increasingly recognizing the potential benefits of algorithmic decision-making in administrative procedures. Two decisions stand out for their main innovative approach:

- *Altomare and others- allocating teachers on short-term contracts*

The State Council decision 2270/2019 in the case of *Altomare and others vs. Ministry of Education*.[333] In this case, the Italian Higher Administrative Court was asked to assess the legitimacy of the adoption of an algorithm for the performance of an administrative activity by the Ministry of Education. The Court ruled that the use of an algorithm by the Ministry for teacher reassignment and recruitment was unlawful due to a lack of transparency. The Court emphasises the necessity for transparency in algorithmic decision-making. This transparency must include all aspects of the algorithm, including its authorship, development process, decision-making criteria, and the data used.[334] Such transparency ensures that the results align with legal requirements and allows for the algorithm to be fully understood and scrutinised, both by citizens and judges.

The legal obligation extends beyond mere transparency; the ***administrative judge must have full access*** to and understanding of the algorithm to effectively evaluate the legitimacy of decisions made through automated processes.[335] This scrutiny includes assessing the logic, reasonableness, and correctness of the algorithm's application. Failure to provide this level of transparency and scrutiny can result in the invalidation of the decision-making process.[336]

In this specific case, the lack of transparency in the algorithm used for assigning teachers to schools led to irrational outcomes. Teachers with more experience were placed in undesired locations far from their residences, while less qualified teachers received their preferred assignments.[337] This lack of transparency and the resulting illogical outcomes violated principles of impartiality and transparency, leading to the annulment of the administrative decision.

- *ADM- arranged teacher's transfers.*

The Italian Council of State addressed the use of algorithms in administrative procedures, particularly in a case involving the transfer of secondary school teachers.[338] In this instance, an algorithm malfunction led to errors in the assignment of teaching locations. While

[333] Consiglio di Stato, 8 Aprile 2019, N. 2270/2017 Reg. Ric. (It) ('Altomare').
[334] Ibid. [para 7-8,2].
[335] Ibid. [para 8,4].
[336] Ibid. [para 9].
[337] See more about this case in Fusco, Carlotta. (2023). The Use of Artificial Intelligence in the Decision-Making Processes of the Public Administration: Regulations and Executive Practice—The Case of the Italian Public Administration. *Innovations and Economic and Social Changes due to Artificial Intelligence: The State of the Art.*
[338] Consiglio di Stato, 8472 Dicembre 2019, N. 8472/2019 Reg. Ric. (It).

acknowledging the benefits of algorithms for efficiency and neutrality, the Court stressed that their use must adhere to core principles to protect individual rights.[339]

The judgement emphasised three key principles derived from EU law. First, it reiterated the importance of transparency, as established in previous decisions. Public administrations must ensure full transparency regarding the algorithm's operation, including its design, decision-making processes, and data handling.[340] This transparency is crucial so that the process and outcomes can be reviewed and understood by those affected.

Secondly, the Court emphasised the principle of non-exclusivity, aligning with Article 22 of the GDPR. This principle dictates that decisions should not be made solely by algorithms. Instead, individuals must have the right to human intervention in the decision-making process to validate or contest the automated outcomes.[341] This safeguard ensures that algorithmic decisions are not absolute and can be subject to human review and adjustment when necessary.

Finally, the Court stressed the principle of non-discrimination. Algorithms must be designed and implemented in a way that prevents discriminatory outcomes based on factors such as race, ethnicity, religion, or other protected characteristics.[342] This requires careful management of data input and close collaboration with algorithm designers to ensure that the system does not perpetuate or exacerbate existing biases.

These decisions marked a turning point, recognizing the benefits of automated decision-making in public administration in Italy. In its rulings, the State Council established that algorithms have legal value in shaping administrative actions and must comply with principles such as transparency, effectiveness, and non-discrimination. Moreover, it emphasised that the source code of these algorithms, once used in public decision-making, should be considered an administrative document subject to disclosure under freedom of information laws. This stance underscores the importance of transparency, even when commercial interests are involved, and highlights the legal obligation for public administrations to ensure the accessibility of such algorithms.

---

[339] Ibid. [para 7,1]. See also Pressi, Matteo. (2021). The Use of Algorithms within Administrative Procedures: National Experiences Compared through the Lens of European Law. *Review of European Administrative Law, 41(69).*

[340] Ibid. [para 13,1].

[341] Ibid. [para. 15,2].

[342] Ibid. [para 15,3]. To expand on this aspect see: Gontarz, Igor. (2023). Judicial Review of Automated Administrative Decision-Making: The Role of Administrative Courts in the Evaluation of Unlawful Regimes. *ELTE Law Journal, 2023(1), 156.*

Netherlands : X (formerly Twitter)- Guano system

The District Court of Amsterdam addressed the practice of 'shadowbanning'[343] by social media platform X (formerly Twitter) in the context of the GDPR.[344] The case started when a user's account was shadowbanned after posting content critical of European child pornography prevention plans. The court determined that X's shadowbanning practice constitutes **automated decision-making** under Article 22(1) of the GDPR. This classification mandates that the platform provider must inform users and provide precise details about the logic behind such actions.

The court highlighted the lack of human intervention in the shadowbanning process, reinforcing its classification as automated decision-making with significant user impact.[345] The ruling underscored X's obligation under GDPR to proactively provide transparent information about shadowbanning, including its logic, significance, and potential impact. Furthermore, the court ordered X to provide access to specific data, including reputation scores, labels, and information processed by their '*Guano*' system, which tracks account actions. The court rejected X's business secrecy defence, reiterating that the access request pertained to the processing of personal data.[346]

This case sets a significant precedent in the interpretation of GDPR requirements for automated decision-making in social media platforms, emphasising the need for transparency and user rights in the digital age. While this case primarily concerns private sector practices, it offers valuable **insights for public administrative bodies** implementing solely automated decision-making systems. The court's ruling underscores the importance of transparency and accountability in automated processes, principles equally or even more applicable to the public sector. Public entities deploying such systems should take note of the requirement to establish clear procedures and make available the logic, significance, and potential impact of their automated decision-making systems. This approach not only aligns with GDPR principles but also promotes trust and accountability in public administration, ensuring citizens are informed and empowered when interacting with automated government services.

4.4 Legal, regulatory and ethical frameworks applicable to decision-making use of AI systems in public administration

4.4.1 Legal framework

Similar to the design of AI and the organisational implementation/embedding of AI in public administration, the legal framework on public-decision making using AI systems is still evolving across international, EU and national levels. In the following sections, we introduce key legal provisions on the use of AI systems and fundamental rights issues arising therefrom

[343] It is a practice that consists of locking (a user) from a social media site or online forum without their knowledge, typically by making their posts and comments no longer visible to other users.
[344] Rechtbank Amsterdam, Case C/13/742407 / HA RK 23-366, ECLI:NL:RBAMS:2024:4019. <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2024:4019>. .
[345] Ibid. [para 4.4-4.7].
[346] Ibid. [para 4.36].

in the context of public administration. We also provide examples of case law from the Court of Justice of the European Union as well as national courts.[347]

a. Transparency and Accountability

Transparency as a necessary tool for rights plays an important role in facilitating accountability. As such, it is instrumental for achieving a potential multitude of other fundamental values which depend on its existence.[348] It is part of EU law as reflected by Article 1 of the Treaty on European Union (TEU) which requires decisions to be taken as openly as possible to the citizens.[349] When this obligation is not complied with, it can generate a plethora of fundamental rights issues such as the violations of the rights to privacy, data protection, non-discrimination and others.[350]

Three provisions of the AI Act are prominent for the purposes of transparency – namely, articles 50, 13, and 86. Article 50 creates a general transparency obligation that is applicable to all-risk AI systems which are intended to interact with natural persons. In those situations, individuals are to be informed that they are interacting with an AI system.[351] Article 13 requires high-risk AI systems to be designed for transparency. This includes providing comprehensive usage instructions, which *inter alia* encompass potential risks to health, safety, or fundamental rights, and ensuring explainability. The AI Act's Article 86 requires AI systems to be inherently explainable, resembling the GDPR's data protection approach, thereby facilitating the exercise of the right to an explanation.

The Law Enforcement Directive (LED) regulates the protection of individuals' data by law enforcement authorities, including automated processing, and mandates a right to access information as a transparency measure in Article 14.[352] The right to access data is crucial for individuals to identify and correct errors in automated decisions.[353] However, this right can be limited in law enforcement contexts to protect investigations, national security, and the rights of others.

---

[347] NB: This list is non-exhaustive.

[348] David Heald, 'Transparency as an Instrumental Value' in Cristopher Hood, David Heald (eds), *Transparency: The Key to Better Governance?* (OUP 2006) 59-73.

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

[350] See examples mentioned in Introduction; Tweede Kamer der Staten-Generaal: Verslag - Parlementaire ondervragingscommissie Kinderopvangtoeslag. Ongekend onrecht. (Kamerstuk 35 510, nr. 2). 2020. (Report of the Inquiry Commission on child benefits payments of the Dutch Parliament).

[351] This obligation does not apply to law enforcement authorities in the course of detecting, preventing, investigating and prosecuting criminal offences. In fact, throughout the AI Act, the transparency elements for the use of AI systems, whether high-risk or general purpose, in the law enforcement sector have been limited; See Article 50 AI Act.

[352] See Article 14 LED.

[353] Diana Dimitrova, Paul De Hert, 'The Right of Access Under the Police Directive: Small Steps Forward' in M Median, A Mitrakas, K Rannenberg, E Schweighofer, N Tsourlas (eds), *Privacy Technologies and Policy: 6th Annual Privacy Forum* (Springer 2018) 111-131.

Transparency has been one of the motivators for the creation of the General Data Protection Regulation (GDPR).[354] As such, it aims to provide accountability and control over personal data.[355] Key provisions for transparency are: Article 5(1)(a) - which establishes the general principle of transparency as one of the core data protection principle; Articles 12-15 further include concrete transparency obligations in the form of access to information by the data subject in an understandable way (Article 12), *ex ante* obligations which include the provision of 'meaningful information about the logic involved' and the envisioned consequences of data processing in the case of automated decision-making (Article 13(2)(f) & 14(2)(g)), and *ex post* possibility for data subjects to access meaningful information and invoke their rights of access (Article 15(1)(h).

The public law framework across the EU generally requires authorities to provide explanations for their decisions as part of their transparency obligations. Although public law varies across EU Member States, the obligation to justify decisions is common in most national legislations, and its effectiveness will depend heavily on a robust transparency framework, especially in the context of AI. The duty to provide reasons strengthens the position of individuals by granting them access to reasoned information on decisions made with the assistance of automated systems. However, this duty is not always adequate when dealing with automated decision-making (ADM), often leaving the courts to assess such matters. So far, only a few courts have addressed these issues. For example, Czech case law has ruled that the obligation to justify decisions remains intact even when automated means are used.[356] Similarly, the Lithuanian Supreme Administrative Court highlighted the principles of legality and good administration in cases involving automated decisions, even when these decisions are subject to appeal.[357]

However, specific public law regulations concerning AI systems in decision-making are still scarce. This is partly due to the prominent role of data protection law in this area and the assumption that general public law principles apply regardless of the decision-making method. Recognizing the limitations of existing public laws, several European countries have enacted legislation with specific provisions on automated decision-making. For instance, Sweden's Administrative Procedure Act allows decisions to be made by officials or automatically[358], while Germany's *Verwaltungsverfahrensgesetz* permits administrative decisions to be issued entirely by automated means where permitted by law and where no

[354] See Recital (13) of the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 (GDPR).

[355] See for instance Recital (39) GDPR.

[356] See Supreme Administrative Court of Czech Republic, Case No. 3 Ads 33/2012 – 30.

[357] See Supreme Administrative Court of Lithuania, Case No. A-275-822/2020.

[358] See Section 28 of the Swedish Administrative Procedure Act (2017:900) available at <https://www.government.se/government-policy/democracy-and-human-rights/the-administrative-procedure-act-2017900/> accessed 5 March 2024; Johan Wolswinkel, 'Comparative Study on Administrative Law and the Use of Artificial Intelligence and Other Algorithmic Systems in Administrative Decision-making in the Member States of the Council of Europe' (Comparative Study for Council of Europe, CDCJ(2022)31, December 2022) 17.

administrative discretion is required.[359] France's Law for a Digital Republic goes further to ensure transparency, with Article R. 311-3-2 of the French Public Relations and Administration Code obligating public authorities to disclose, without violating legally protected secrets, the degree and mode of algorithmic processing involved in decision-making, the sources and data processed, the parameters and their weighting applied to the individual's situation, and the operations performed by the algorithm.[360] In Portugal, the Charter of Human Rights in the Digital Age, established by Law No. 27/2021 of May 17, 2021, regulates AI and robots, with Article 9 mandating that high-impact AI-assisted decisions be appealable and auditable according to the law.[361] Lithuania includes access to source code as part of transparency.[362] In France, source code is considered an administrative document, making it accessible to citizens upon request.[363] Other countries either lack specific legislation on AI-assisted decision-making or are in the process of developing such laws. This highlights the significant challenges in extending administrative principles and remedies to AI used in decision-making.

I. EU case law

CJEU: Ligue des droits humains ASBL v Conseil des ministres: Request for a preliminary ruling from the Cour constitutionnelle

In this case the CJEU addressed critical issues surrounding the use of automated systems for processing Passenger Name Record (PNR) data. The Court's decision was introduced in the previous stages but this case also highlights several important aspects within the decision making stage.[364]

The case addresses the interpretation and validity of the EU's Directive 2016/681 on the use of passenger name record (PNR) data for counterterrorism and serious crime, in light of the Charter of Fundamental Rights of the European Union (CFREU). The court emphasised the need for non-discriminatory, reliable, and specific criteria when analysing PNR data, and highlighted the importance of both incriminating and exonerating factors. The judgement

[359] Art. 35(a) Verwaltungsverfahrensgesetz (VwVfG) available at <https://www-gesetze--im--internet-de.translate.goog/vwvfg/BJNR012530976.html?_x_tr_sl=de&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=sc> accessed 5 March 2024.
[360] Act No 2016-1321 of 7 October 2016 (France).
[361] See Law No. 27/2021 at <https://files.diariodarepublica.pt/1s/2021/05/09500/0000500010.pdf> accessed 6 March 2024.
[362] Johan Wolswinkel, 'Comparative Study on Administrative Law and the Use of Artificial Intelligence and Other Algorithmic Systems in Administrative Decision-making in the Member States of the Council of Europe' (Comparative Study for Council of Europe, CDCJ(2022)31, December 2022) 19.
[363] See Article 6 of the Law for a Digital Republic which requires public authorities to 'publish the rules defining the major algorithmic processes used in the realisation of their tasks when they are used for individual decisions'; Cyrille Chausson, 'France opens the source code of tax and benefits calculators to increase transparency' (joinup.ec.europa.eu, 20 November 2016) <https://joinup.ec.europa.eu/collection/egovernment/document/france-opens-source-code-tax-and-benefits-calculators-increase-transparency> accessed 6 March 2024.
[364] Case C-817/19 *Ligue des droits humains ASBL v Conseil des ministres* [2022] ECLI 491.

also scrutinises the use of AI in automated data analysis, emphasising that such systems must not operate without **human oversight** due to the potential risks of opacity and lack of transparency. The court insists on the necessity for individuals to understand the criteria and workings of AI systems to ensure their right to an effective remedy.

II. National case law

Italy: ADM and human rights, decision of the Supreme Court (Corte di Cassazione)

Italy's Supreme Court of Cassation ruled that consent is invalid if an individual is subjected to an automated decision-making (ADM) system that could impact their rights without sufficient understanding of the system's underlying logic.[365] The *Associazione* contended that individuals consented to the ADM by joining its platform and uploading documents voluntarily. However, the Court determined that simply joining the platform does not constitute acceptance of an ADM system that evaluates individuals based on personal data unless they are fully informed of the system's logic and key algorithmic elements. Therefore this case highlights the transparency and accountability measures not only in the GDPR but also in the different national law provisions.

Denmark: Danish Data Protection Agency in regards to the Tilsyn med Udbetaling

Udbetaling Danmark is a public authority responsible for managing various public benefits - employing standardised templates to respond to access requests, which indicated that automated decisions were rendered in specific instances.[366] However, automated decision-making was not applied uniformly to all benefit types, being confined to income-based benefits such as housing allowance. Despite this, the same template was used for all access requests, leading to a lack of specific information about automated decisions for certain data subjects and thereby violating Article 15(1)(h) of the GDPR. The Danish DPA determined that Udbetaling Danmark neglected to inform data subjects about the existence of automated decisions concerning the granting of specific benefits, in contravention of Article 15(1)(h) of the GDPR which requires transparency.

Sweden: Kammarrätten i Göteborg

When confronted with a request under freedom of information legislation, the Swedish municipality of Trelleborg only provided '136,000 lines of rules across 127 XML files' on their welfare software without offering any explanation or access to the underlying data.[367] This lack of clarity made it impossible to fully understand the rules, effectively maintaining the system as a 'black box'. This example highlights that true transparency requires not just

[365] Supreme Court (Corte di Cassazione), Cass. Civ. Ord. Sez. 1 Num. 14381, 25 May 2021.

[366] Datatilsynet, Supervision of Udbetaling Danmark's processing of personal data, 26 February 2020, available at https://www.datatilsynet.dk/afgoerelser/afgoerelser/2020/feb/tilsyn-med-udbetaling-danmarks-behandling-af-personoplysninger.

[367] Katarina Lind, 'Central authorities slow to react as Sweden's cities embrace automation of welfare management' (*AlgorithmWatch*, 17 March 2020) <https://algorithmwatch.org/en/trelleborg-sweden-algorithm/> accessed 17 April 2024.

the disclosure of code, but also sufficient information and explanations for the audience to accurately reconstruct the algorithm.[368]

b. Bias and Discrimination

The right to non-discrimination, or equal treatment, is a cornerstone of Union Law and one of the European Union's foundational values, as outlined in Article 2 of the Treaty on European Union (TEU). It also poses as a fundamental right, codified in Article 21 of the EU Charter of Fundamental Rights. Naturally, this essential right has been integrated into the AI Act to ensure that, as technology rapidly advances, these principles are upheld.

Under the AI Act, high-risk AI systems are subject to stringent requirements before they can enter the market, including the need for a **high quality of the datasets feeding the system to minimise risks and discriminatory outcomes**.[369] Furthermore, uniform standards for all high-risk AI systems should align with the Charter of Fundamental Rights of the European Union and be non-discriminatory.[370] Among other things, technical inaccuracies in AI systems designed for remote biometric identification can result in biassed outcomes and discriminatory effects; this is exactly why the use of predictive policy algorithms in law enforcement and public safety has sparked significant concerns.[371]

While these advanced systems have been praised for their potential to enhance security and efficiency, growing evidence shows they may disproportionately target minority communities, leading to increased surveillance, over-policing, and potential violations of civil liberties. These issues stem not only from historical biases embedded in the data used to train these algorithms but also from the way these tools are implemented, which can perpetuate systemic inequalities.

Moreover, individuals applying for or receiving public assistance are often in vulnerable positions, heavily reliant on these benefits and services. If AI systems are used to determine whether such benefits should be denied, reduced, revoked, or reclaimed, the consequences could be severe, potentially infringing on their fundamental rights, including the right to non-discrimination. Therefore, it is imperative that the deployment of AI in public administration is handled with extreme care, ensuring that these systems do not exacerbate existing inequalities but instead contribute to a more just and equitable society.

To safeguard these rights effectively, continuous monitoring, transparency, and accountability must be prioritised in the deployment of AI systems. Public entities need to ensure that AI-driven decisions are not only fair and unbiased but also explainable to those affected, providing mechanisms for redress and ensuring that AI serves as a tool for enhancing, rather than diminishing, civil liberties.

---

[368] Sweden, Kammarrätten i Göteborg, 2020-02-21, Mål nr 6973-19
[369] AI act, recital 38.
[370] AI act, recital 13
[371] AI act, recital 33.

I. National case law

Slovakia: eKasa case- Tax authority

The extensive analysis on this case can be found within the organisational stage case law of this handbook. However, for the purposes of this stage the case in sum is about the Slovak tax authority employed three distinct algorithms: a chatbot 'TAXANA', an enterprise resource planning (ERP) system 'eKasa', and a risk-scoring machine-learning algorithm 'AIS-R'. The controversy centred around the 'eKasa' system and its associated 'AIS-R' algorithm, which processes data to predict VAT fraud risks. The Slovak Constitutional Court ruled that transparency obligations must take precedence over trade secrets in AI-assisted decision-making.[372] The court also emphasised the importance of individual access to information about automated assessments for the right to effective remedy.[373]

The Court stressed the necessity for additional legal protections against such automated assessments by State agencies, including ensuring the accuracy and fairness of the criteria used, informing individuals about the assessment process, monitoring the system's effectiveness, and offering ways for individuals to contest errors or biases within the system.

c. Privacy and Data Protection (not extensive)

I. EU case law

CJEU: Joined Cases C-511/18, C-512/18 and C-520/18 La Quadrature du Net and Others[374]

The judgement originates from requests for a preliminary ruling on the interpretation of Article 15(1) of Directive 2002/58/EC regarding the processing of personal data in the electronic communications sector, analysed in the context of the Charter of Fundamental Rights of the European Union and Article 4(2) TEU.

The Court considered five key questions: i) Whether national legislation that requires electronic communication providers to forward data to security agencies falls within the scope of Directive 2002/58. ii) Whether Article 15(1) of Directive 2002/58 prohibits national laws that mandate general and indiscriminate retention of traffic and location data. iii) Whether national laws allowing automated analysis and real-time collection of traffic and location data are compatible with Directive 2002/58. iv) Whether Directive 2000/31 precludes national laws requiring the general and indiscriminate retention of personal data by online service providers. v) Whether national courts can apply laws requiring general data retention despite conflicts with EU law.

[372] Constitutional Court of the Slovak Republic PL., Judgement 492/2021 Z.z. of the ÚS 25/2019-117 (2021) (hereinafter 'eKasa'). <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2021/492/20211217>.
[373] See eKasa judgement, para 133.
[374] Joined Cases C-511/18, C-512/18 and C-520/18 *La Quadrature du Net and Others* [2020] OJ C 433/3.

The Court ruled that Directive 2002/58 applies to national legislation involving the **collection and retention of personal data by electronic communication providers.** The CJEU rejected the argument that such activities are purely state functions outside the Directive's scope, as they involve data processing by service providers, which is regulated by EU law. The Court emphasised that any exceptions to the confidentiality of electronic communications under Article 15(1) of the Directive must comply with EU law principles, including proportionality and fundamental rights. The Court found that general and **indiscriminate data retention by providers is not allowed**, as it violates the principles of privacy and freedom of expression.

The Court highlighted the risk of profiling from the retention of traffic and location data, which can lead to significant invasions of privacy. It concluded that data retention for national security must be strictly proportionate to the threat and subject to effective oversight. The CJEU also ruled that **automated data analysis and real-time data collection for anti-terrorism purposes must be narrowly tailored and based on genuine, foreseeable threats**, with limited retention periods and prior judicial or independent administrative review.

Finally, the Court held that national courts cannot limit the temporal effects of declarations of illegality under national law if they conflict with EU law. However, national criminal courts must disregard evidence obtained through illegal data retention if it prevents suspects from effectively contesting the information.

II. National case law

Germany: Mobile Phone Data Extraction in Asylum Procedures

On February 16, 2023, the German Federal Administrative Court (BVerwG) declared the Federal Office for Migration and Refugees' (BAMF) routine practice of analysing data carriers, including mobile phones, during the registration of asylum applicants as unlawful.[375] This ruling, following advocacy by a non-profit organisation, represents a significant advancement in protecting asylum seekers' data privacy rights. The Court emphasised the necessity and proportionality of mobile phone data analysis, stipulating that it should only be employed when less intrusive methods prove ineffective.

[375] BVerwG 1 C 19.2 [2023] <https://www.bverwg.de/160223U1C19.21.0> accessed 3 September 2024.

The case in question involved an Afghan asylum seeker whose mobile phone was analysed despite her providing other documents to establish her identity. The court found that the BAMF's routine practice of analysing mobile phones without prior necessity assessments was illegal. This judgement marks a significant victory for asylum seekers' data protection and privacy rights. The BAMF will now have to assess whether less intrusive methods are available before analysing mobile phone data. However, the ruling leaves some issues unresolved, particularly concerning the compatibility of this practice with EU data protection laws, such as the GDPR. The case also highlights broader concerns about the proportionality and suitability of using mobile phone data in asylum procedures, given the potential unreliability of the data and the lack of transparency in the process. The legal challenge against the BAMF continues, with further scrutiny expected from the Federal Data Protection Commissioner.

Sweden, Kammarrätten i Göteborg, 2020-02-21, Mål nr 6973-19

A reporter requested the source code behind the software that Trelleborg municipality uses for automated decision-making for livelihood support.[376] The municipality rejected the reporter's request on the grounds that the source code was not a document, but a product. The reporter appealed the decision to the Court of Appeal, arguing that the software would be released. The municipality stated that they were the owners of the source code and that it paid large sums for it. According to the municipality the code was covered by confidentiality. The Court of Appeal initially stated that the software was a public document of the municipality. The court then stated that confidentiality rules apply to information about an individual's business or operating conditions when he/she entered into a business relationship with an authority, if for a special reason it can be assumed that the individual suffers damage if the information is disclosed. It prevails according to the court therefore has a particularly strong presumption for publicity. In that the municipality owns the software, according to the court, it has not emerged something that makes secrecy prevent disclosure.

d) Fairness and Due Process

I. National case law

The Netherlands, Dutch Child Benefit Scandal,[377] Data Protection Authority, 25 November 2021

The Dutch Data Protection Authority (AP) levied a €2,750,000 fine against the Dutch Tax Administration for unlawfully processing personal data without proper legal grounds and

[376] Kammarrätten i Göteborg, dom 2020-02-21 i mål nr 6973–19.
[377] Autoriteit Persoonsgegevens Decision <https://autoriteitpersoonsgegevens.nl/uploads/imported/boetebesluit_belastingdienst.pdf> accessed 3 September 2024.

violating the principle of fairness, thereby breaching Article 5(1)(a) in conjunction with Article 6(1)(e) of the GDPR, and Article 6 in conjunction with Article 8 of the Personal Data Protection Act (Wbp). From 2013 to 2019, approximately 26,000 parents were wrongfully accused of fraudulently claiming childcare benefits, resulting in demands for full repayment. The AP identified three unlawful data processing incidents, two of which also breached the fairness principle. Due to the gravity, extended duration, and clear responsibility for these violations, the AP imposed a significant administrative fine under Article 58(2) and Article 83 of the GDPR.

e) Human Oversight

While the primary goal of advancing these technologies and introducing artificial intelligence is to reduce human involvement in operations and enhance efficiency, a paradox arises. Ironically, due to their potential effects on fundamental rights, these systems require strict oversight to prevent any unintended consequences. As such, it is crucial that these AI systems are designed and developed with mechanisms that allow natural persons to oversee their functioning effectively.

The AI Act underscores the necessity of incorporating appropriate human oversight measures by the provider before the system is placed on the market or put into service; this ensures that human operators can intervene if needed, maintaining control over the AI's actions.[378]

In alignment with this, Article 14 of the AI Act is dedicated to establishing the framework for such oversight, reinforcing the European Commission's commitment to a human-centric approach in AI deployment.[379] As such, the requirement for human control or interference embedded within Article 14(1) represents a critical safeguard, aimed at compelling designers of high-risk AI systems to include human oversight functions. Virtually, this safeguard is not only a technical requirement, but also a fundamental necessity to mitigate the risks associated with AI malfunctions.

Nevertheless, it is contended that this provision is fraught with impracticalities.[380] Humans are often unable to fully grasp the capabilities and limitations of an AI system, and since the article lacks specifics on when and how human oversight should be implemented, it renders the oversight ineffective. Moreover, it neglects to define what human oversight should focus on, such as ensuring non-discrimination and fairness.

---

[378] Recital 48 of the EU AI Act.

[379] https://www.euaiact.com/key-issue/4

[380] Ebers, M.; Hoch, V.R.S.; Rosenkranz, F.; Ruschemeier, H.; Steinrötter, B. The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). J 2021, 4, 589-603.

I. National case law

Denmark: Danish Data Protection Agency in regards to the Tilsyn med Udbetaling

Udbetaling Danmark is a public authority responsible for managing various public benefits - employing standardised templates to respond to access requests, which indicated that automated decisions were rendered in specific instances.[381] However, automated decision-making was not applied uniformly to all benefit types, being confined to income-based benefits such as housing allowance. Despite this, the same template was used for all access requests, leading to a lack of specific information about automated decisions for certain data subjects and thereby violating Article 15(1)(h) of the GDPR. The Danish DPA determined that Udbetaling Danmark neglected to inform data subjects about the existence of automated decisions concerning the granting of specific benefits, in contravention of Article 15(1)(h) of the GDPR which requires transparency.

II. Ongoing cases

Spain, National Court of Justice, BOSCO Case,[382] 30th April 2024

In 2018, the *Civio Foundation* identified that numerous eligible individuals were not receiving the 'bono social,' a government subsidy for electricity bills, due to errors in BOSCO, the algorithm used to determine eligibility. After requesting access to BOSCO's source code and receiving only partial information, Civio demonstrated that the algorithm contained significant flaws. However, access to the complete source code was denied on the grounds of intellectual property rights. The Civio Foundation subsequently pursued legal action, but both the initial ruling and the appeal upheld the denial of access. Civio has now filed a preparatory appeal against the National Court decision with the Spanish Supreme Court, marking the first case in Spain to address the issue of algorithmic transparency.

Case T-37/24, CB v Commission, 22 January 2024

The applicant asserted claims of insufficient transparency regarding the assessment methodologies, the unlawful evaluation of the applicant by means of artificial intelligence, a deficiency in legal certainty, and an inadequate provision of justifications.

---

[381] Datatilsynet, Supervision of Udbetaling Danmark's processing of personal data, 26 February 2020, available at https://www.datatilsynet.dk/afgoerelser/afgoerelser/2020/feb/tilsyn-med-udbetaling-danmarks-behandling-af-personoplysninger.

[382] Civio, 'The Spanish National Court again rejects opening the source code validating electricity subsidy applications' (May 2024) <https://civio.es/novedades/2024/05/08/spanish-national-court-rejects-opening-source-code-for-electricity-subsidy-applications/> accessed 3 September 2024.

4.4.2 Ethical concerns relating to the use of AI systems in public decision-making

Public authorities increasingly rely on algorithmic systems to inform or take administrative acts. Broadly speaking, these acts include individual and normative legal acts, physical actions of the administration in pursuit of its public authority, and refusal or an omission to act.[383]

It is through administrative acts that public administration intervenes to regulate the behaviour of people and legal entities in their relationships with the state. These decision-making processes are increasingly being delegated to algorithmic systems which creates a parallel layer to traditional governance that is now called algorithmic regulation. Algorithmic regulation refers to the use by public administration of algorithmic systems to inform and adopt administrative acts.[384]

Algorithmic regulation systems are best seen as a socio-technical infrastructure. The deployment of this infrastructure requires an assessment of both the technical and the social forces behind technology design.[385] There are several key societal risks related to algorithmic regulation systems: risk of human error, bias, opacity, dependence on data and proxies, and human agency.[386]

The **human error** risk can manifest itself throughout the algorithmic regulation system's lifecycle and can be the result of programming errors, inaccurate data models, or wrongful deductions based on seemingly causal relationships between phenomena. This risk can lead, for example, to allocation harms, ill-informed policy decisions, or indeed scenarios with disastrous, long-term consequences.[387]

The **risk of bias** is manifested predominantly in data-driven algorithmic regulation systems. It can transpire in the choice of historical data which may be skewed towards one or more groups in the society or that may contain wrongful assumptions about the characteristics of one or more groups of people. The realisation of this risk can have significant consequences when enforced on a scale through the socio-technical infrastructure that are algorithmic regulations systems.

---

[383] As defined by Nathalie Smuha in The Algorithmic Leviathan and the Rule Of Law: An examination of how algorithmic regulation in the public sector can lead to 'algorithmic rule by law', and a critical evaluation of EU legal safeguards, PhD Thesis, KU Leuven, December 2022 (unpublished), at 4., referring to Committee of Ministers of the Council of Europe, 'Recommendation Rec(2004)20 of the Committee of Ministers to Member States on Judicial Review of Administrative Acts' 2.
[384] Smuha, Op cit, 5.
[385] Smuha, Op. cit, 64.
[386] Based on Smuha, Op cit., 69-87.
[387] Such as the UK Post Office's Horizon system deployed to perform transaction, accounting and stocktaking tasks. See on the legal and ethical issues the statement of BCS, The Chartered Institute for IT, https://www.bcs.org/articles-opinion-and-research/bcs-statement-on-the-post-office-horizon-it-scandal/.

The **risk of opacity** is related to algorithmic regulation systems' deficit of transparency which is manifested in the fact that the system is used, the way in which it is used, and the way in which it works.[388] For example, when administrative acts concerning citizens are delegated to an algorithmic regulation system, that choice implicates the responsibility of the officer, as a minimum, to inform the citizen that they are subject to algorithmic regulation. The way in which a system is used could also run into the risk of opacity because of the implicit choices made by the system designers and developers which are not necessarily communicated to the deployers or users, if at all. Finally, the risk of opacity can also be present in the way the system works, which refers to the traditional issue of the system as a black box and the explainability of algorithmic decision making.

The risk of **dependency on data and proxies** concerns the reliance on data to derive a model used to make decisions. In these cases, public authorities should be aware of the gap between the data and the resulting model. The risk of reducing the individual human being to numerical utility function is particularly pronounced in data-driven algorithmic regulation systems. This is so not least because these systems could interfere with both personal autonomy and individual privacy.

The impact on personal autonomy of individuals is directly related to the **risk of losing human agency**, that is, the ability of humans to act autonomously and intentionally. This risk has a direct impact on the moral responsibility of public authority decision makers.

The deployment of systems for algorithmic regulation is usually triggered by governments and policymakers' desire to promote public safety, social stability or continued innovation.[389] However, the deployment of these systems comes with **increased ethical responsibility** on the part of public administration. First and foremost, this implies that algorithmic regulation systems should not be seen as responsible entities. Instead, the people and organisations involved in their development and deployment must take responsibility and be accountable for their acts and omissions to act.

Algorithmic regulation systems are **manufactured artefacts**, ie, they are built using conventional tools of software development and exist in the socio-technical system that we call society. At the very least, these systems are limited by the laws of physics which apply to any computational model such as the Universal Turing Machine, in particular, the limitations of memory (storage) and time (execution time).

This implies that algorithmic regulation systems cannot be decoupled from this system and any ethical governance or guidance of such systems should start from the premise that these systems are no moral agents. Indeed, algorithmic applications and systems have no agency. As such, public administration cannot and should not expect that ethical conduct,

[388] Smuha, Op. cit, 75.
[389] Theodorou, A., Dignum, V. Towards ethical and socio-legal governance in AI. Nat Mach Intell 2, 10–12 (2020). https://doi-org.kuleuven.e-bronnen.be/10.1038/s42256-019-0136-y.

trustworthiness or responsibility are categories that can be applied to these material artefacts. Ethical deployment and use of algorithmic regulation systems is not about attributing some sort of responsibility to machines; it is about making people and entities more responsible and conscious of their choices when deploying such systems.[390]

High-level ethics guidelines often fail to address **cultural differences**. Policies should consider the cultural variety between societies affected by algorithmic regulation systems. Recognising and understanding the cultural contexts in which these systems operate is essential to ensure that they are fair, inclusive, and effective. It is the elusive nature of these terms, as used by various ethics guidelines, that creates a considerable risk of failure of public regulation of AI.[391] This means adapting these categories to local norms, values, and societal expectations, which may vary significantly from one region to another. A one-size-fits-all approach may not be appropriate, as it could lead to misunderstandings, resistance, or even harm. Therefore, it is essential to engage with local stakeholders, including community leaders, civil society organisations, and end-users, to co-create ethical frameworks that are culturally sensitive and contextually relevant. This collaborative approach can help bridge the gap between global ethical standards and local realities, fostering greater acceptance and trust in algorithmic regulation systems.

Most of the algorithmic regulation systems that are in use or under development are in fact developed by private companies. Public authorities often lack the technical and financial resources to develop, deploy and maintain such systems at scale in pursuit of their objectives of cost reduction while improving public welfare. The process of engaging in closer collaboration with the private sectors runs the risk of **outsourcing ethical decisions** too. Indeed, ethical decisions may be made by technology developers or within the private confines of corporate boardrooms, thereby gaining the enforceability of technological affordances that embed them.[392] The different ordering of values in systems will lead to different outcomes in the same situation.[393] The ethical responsibility of both public authorities and the private companies they collaborate with requires that algorithmic regulation systems clearly identify which ethical theories or considerations are applied in different contexts.

These risks cannot be mitigated by the adoption of ethics-first approaches or by relying solely on the goodwill of people and organisations. This is where ethics should be complemented by **standards and legislation**. Moving beyond ethical guidelines to create technical standards and legislation is crucial to ensure safety, security, and reliability.

For instance, technical standards can provide specific criteria for the development and deployment of algorithmic systems, ensuring that they operate within defined safety and

[390] Ibid.
[391] Ibid.
[392] Hildebrandt, M. *Law for Computer Scientists,* https://lawforcomputerscientists.pubpub.org/pub/nx5zv2ux.
[393] Theodorou and Dignum, Op cit.

performance parameters. Standards can cover a range of aspects, from data quality and algorithmic transparency to user privacy and security. By establishing clear benchmarks, standards can help prevent the misuse of these systems and promote best practices across different industries and sectors.

Legislation, on the other hand, can enforce compliance with these standards and hold organisations accountable for any breaches. Laws mandate regular **audits**, impact assessments, and transparency requirements, ensuring that public authorities and private companies alike adhere to both legal and technical norms of interoperability, safety, and robustness. Law also offers **recourse** for individuals and communities affected by algorithmic decisions, providing mechanisms for redress and accountability.

International **cooperation** is essential in developing these standards and legislative measures. As algorithmic regulation systems increasingly operate across borders, it is crucial to harmonise regulations and foster collaboration among different countries and regions. Ethical frameworks can be the basis for **alignment of the views and understanding the differences** in such discussions. This can help create a coherent and effective global framework for algorithmic governance, reducing the risks of fragmentation and inconsistency.

In addition to technical and legal measures, public authorities and private companies should invest in **education and training**. Computer science curricula should feature more prominently discussions on the societal implications of system design.[394] This includes interdisciplinary courses that cover legal and moral accountability, ensuring that developers understand the broader context and impact of their work.

4.5 Concluding remarks: Decision-making stage

In the decision-making stage of the AI lifecycle, the continuous operation of AI tools within an organisation's processes requires both technological and governance considerations to ensure effective and ethical outcomes. *Technologically,* this stage involves evaluating AI models to assess their real-world performance and fairness, using metrics like ROC curves, AUC, and model calibration across diverse populations. This evaluation is essential for ensuring that AI systems integrate well within organisational workflows and maintain fairness.

From a governance perspective, transparency, documentation, and accountability are crucial. AI systems must be deployed with a clear understanding of their real-world impact, and public servants must interpret their outputs correctly. Transparency is vital for public trust, particularly when AI systems are used in public administration. Organisations must

[394] Theodorou and Dignum, Op cit.

document all stages of model development and make these records accessible for audits and public scrutiny to ensure fairness and compliance with legal and ethical standards.

Moreover, organisations must safeguard individuals' rights by providing mechanisms for challenging AI decisions, seeking explanations, and ensuring human oversight. For example, in public health or welfare scenarios, transparency about decision-making criteria and the ability to appeal decisions are essential for maintaining public trust and protecting fundamental rights.

Real-world case studies, such as the SCHUFA case in Germany or the PNR data processing in France, highlight the importance of transparency, human oversight, and rigorous documentation in AI decision-making. These cases show how AI systems can significantly impact individuals' rights and the importance of establishing robust governance structures to prevent discriminatory or unfair outcomes.

4.5.1 Guiding questions

The aim of this handbook is to equip participants with practical information and real-world case studies to help them mitigate potential conflicts with fundamental rights during AI system development. In this section, we highlight some of the most critical questions that public entities should consider during the decision making stage, ensuring that their AI initiatives are both effective and aligned with legal and ethical standards.

Fundamental rights

- What are the potential risks associated with the decision-making process at each stage of the system's lifecycle?
- How do these risks interconnect and influence one another across different stages?
- Are the processes clear that allow citizens to understand how decisions are made, seek redress if necessary, and ensure that their rights are protected?
- Is there an adequate human oversight procedure in place?
- Does the system and how it is used guarantee fairness and due process?
- What are the possible discriminatory outcomes of the decision-making process, both direct and indirect?
- Are there any risks of opacity in how decisions are made or communicated to affected individuals?
- How can we ensure that decisions are transparent and understandable to those impacted?
- What mechanisms are in place to monitor and evaluate the decision-making process for ongoing accountability?

Due Process and Human Oversight

- Is there an adequate human-in-the-loop mechanism in place during the decision-making process?
- How can we ensure that decisions are not perceived as arbitrary or unjust due to the lack of human oversight?
- What measures are in place to allow individuals to understand or contest decisions made by the system?
- How can we improve transparency in algorithmic decision-making to uphold the right to due process?

Consultation and Expertise

- Who should be consulted to identify and mitigate these risks (e.g., employees, third-party experts, civil society, affected individuals)?
- How can we involve technical, legal, and organisational experts to assess potential risks?
- What role should civil society organisations and potential victims play in the consultation process?
- How can we gather concerns and experiences from a diverse range of stakeholders to inform the decision-making process?
- What methods (e.g., public hearings, expert testimony) can be used to engage stakeholders effectively?

Documentation and Review

- What relevant documents (e.g., impact assessments, protocols, reports) should be reviewed to identify risks?
- How should these documents be analysed to uncover potential risks to fundamental rights?
- What correlations between lifecycle stages and effects should be considered when reviewing these documents?
- How should we document the decision-making process to ensure transparency and accountability?

5. Conclusions

Lorena Flórez, Tobias Nowak

The responsible use of AI in public administration is a complex challenge that requires a well-structured approach throughout the entire AI lifecycle. From design to embedding and decision-making, each stage carries its own set of risks, opportunities, and ethical considerations that must be carefully managed to ensure that AI serves the public interest while protecting fundamental rights.

In the **design phase**, public entities must prioritise a clear understanding of the problems they aim to address before opting for AI as a solution. By thoroughly assessing the appropriateness of AI through processes like technology impact assessments, organisations can avoid the pitfalls of misaligned solutions. Careful attention to data selection, algorithm design, and regulatory compliance is essential to mitigate risks of bias, discrimination, and rights violations. Compliance with legal frameworks such as GDPR and the AI Act is crucial, ensuring that AI systems are not only effective but also ethical.

As AI is **embedded** within public institutions, transparency, accountability, and governance become essential. AI systems must be seamlessly integrated into existing infrastructures, but more importantly, public organisations must establish clear mechanisms to explain how AI-driven decisions are made. Transparency builds trust, while accountability frameworks allow citizens to challenge AI decisions and seek redress when necessary. Ongoing monitoring, oversight, and continuous improvement of AI systems help adapt to evolving societal and technological landscapes, maintaining alignment with fundamental rights and democratic values.

In the **decision-making phase**, public bodies must ensure that AI systems operate fairly and transparently. By thoroughly evaluating AI models and maintaining robust documentation, public organisations can uphold accountability and provide clear explanations of how AI-driven decisions are made. It is essential to offer mechanisms for individuals to contest decisions, appeal outcomes, and ensure human oversight, particularly in sensitive areas like public health or welfare.

Overall, AI in public administration holds great potential to improve efficiency and decision-making processes, but it must be deployed responsibly. Public trust, legal compliance, and ethical governance are the cornerstones of successful AI implementation. This Handbook addressed some of the challenges that the use of AI in public administration creates for the protection of fundamental rights. With the help of existing cases and legal frameworks, it explored issues of transparency, liability, accountability, and the right to privacy and to effective remedies that may arise when using AI to help provide public services, enforce laws and policies and for internal management. We looked at the different challenges found in the different phases of the life cycle of an AI system - how it is designed, how it is integrated into the organisation, and how it is used in decision making. We formulated - with no claim to completeness - a number of guidelines we deem useful for judges, lawyers, and other legal practitioners, on how to approach each phase from a fundamental rights

perspective. We hope to expand on these with what we learn from the transnational training sessions in which we will analyse hypothetical cases based on cases contained in this Handbook. Recounting the guidelines:

For the **design phase**, in which choices are made about the training data, the complexity of the algorithmic system, predictive indicators, labels, accuracy, and transparency:

Fundamental Rights

- Which fundamental rights, such as non-discrimination, equal treatment, privacy, data protection, and human dignity, are potentially affected by this system?
- Could the design of the algorithm unintentionally embed biases?
- Are there discriminatory risk indicators or criteria that could lead to direct or indirect discrimination, especially for vulnerable groups like women with children?
- Does the algorithm's design respect human dignity, or does it reduce individuals to mere data points or scores?
- How does the system ensure fair treatment of all individuals, regardless of their personal circumstances?

Transparency

- How transparent/accessible/public are the criteria, weighting, and cut-offs used in the algorithm?
- Is the system design clear and understandable to ensure that individuals have access to effective remedies if their rights are infringed?

Impact Assessments:

- Does the system involve the collection and processing of sensitive data categories?
- Have comprehensive Data Protection Impact Assessments (DPIA) and Privacy Impact Assessments (PIA) been conducted to identify and mitigate privacy risks?
- Has a thorough Fundamental Rights and Technology Impact Assessment (FRIA) been conducted to evaluate potential adverse impacts on citizens and their fundamental rights?

For the **organisational embedding phase**, in which algorithmic system is embedded and configured within the working process of an organisation, how one interacts with the system:

Fundamental Rights

- Which fundamental rights, such as non-discrimination, equal treatment, privacy, data protection, and human dignity, could be impacted by the AI system?
- How might the system inadvertently perpetuate or amplify existing biases, and what safeguards are in place to prevent this?

- Does the organisation have mechanisms for ongoing monitoring and review of the algorithm's outcomes to ensure alignment with fundamental rights?
- What procedures are in place to allow citizens to challenge AI-driven decisions and seek redress?

Organization's Values and Culture

- Is the AI system interoperable with existing organisational infrastructure, scalable to meet future demands, and aligned with the organisation's core values?
- In what ways might the organisational culture influence the AI system's recommendations and outcomes?
- How is personnel training structured to address and mitigate ethical risks and systemic biases associated with AI use?
- What protocols and procedures are established to guide manual reviews of AI decisions, ensuring human oversight?
- Are diverse stakeholders actively involved in the ongoing review, monitoring, and governance of the AI system to ensure comprehensive oversight?

For the **decision making phase**, in which the output of an algorithmic system is used to come to a decision following the manual review:

Fundamental rights

- What are the potential risks associated with the decision-making process at each stage of the system's lifecycle?
- How do these risks interconnect and influence one another across different stages?
- Are the processes clear that allow citizens to understand how decisions are made, seek redress if necessary, and ensure that their rights are protected?
- Is there an adequate human oversight procedure in place?
- Does the system and how it is used guarantee fairness and due process?
- What are the possible discriminatory outcomes of the decision-making process, both direct and indirect?
- Are there any risks of opacity in how decisions are made or communicated to affected individuals?
- How can we ensure that decisions are transparent and understandable to those impacted?
- What mechanisms are in place to monitor and evaluate the decision-making process for ongoing accountability?

Due Process and Human Oversight

- Is there an adequate human-in-the-loop mechanism in place during the decision-making process?

- How can we ensure that decisions are not perceived as arbitrary or unjust due to the lack of human oversight?
- What measures are in place to allow individuals to understand or contest decisions made by the system?
- How can we improve transparency in algorithmic decision-making to uphold the right to due process?

Consultation and Expertise

- Who should be consulted to identify and mitigate these risks (e.g., employees, third-party experts, civil society, affected individuals)?
- How can we involve technical, legal, and organisational experts to assess potential risks?
- What role should civil society organisations and potential victims play in the consultation process?
- How can we gather concerns and experiences from a diverse range of stakeholders to inform the decision-making process?
- What methods (e.g., public hearings, expert testimony) can be used to engage stakeholders effectively?

Documentation and Review

- What relevant documents (e.g., impact assessments, protocols, reports) should be reviewed to identify risks?
- How should these documents be analysed to uncover potential risks to fundamental rights?
- What correlations between lifecycle stages and effects should be considered when reviewing these documents?
- How should we document the decision-making process to ensure transparency and accountability?

More generally, we formulated questions that are intended to serve as a guide during the training, providing direction and focus for our discussions. Please review them in preparation for the workshop:

- Which legal obligations do public bodies have with regard to their use of algorithms or AI?
- In particular: How do public bodies using algorithms or AI have to comply with the citizens' right to transparency?
- Which rights / remedies / claims do individuals and other private parties have vis-à-vis public bodies that use algorithms or AI?
- How will the new EU AI Act change the situation in your country?
- What can judges do or not do in cases concerning the use of algorithms

or AI by the public administration?

- How can judges access the necessary socio-technical expertise to make a decision in such cases?
- Concerning biases in (the use of) algorithms or AI systems by the public administration: What about the difference between bias in design and bias in application?