EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

cepej
European Commission for the Efficiency of Justice
Commission européenne pour l'efficacité de la justice

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

EDPS 20

Wojciech Rafał WIEWIÓROWSKI

Supervisor

Human Oversight of the automated decision making process

Council of Europe European Commission for the efficiency of justice (CEPEJ)
European Cyberjustice Network, November 21th, 2025

Dr hab. Wojciech R. Wiewiórowski,

Departament of Legal Informatics, University of Gdansk (Poland)
European Data Protection Supervisor (EDPS)

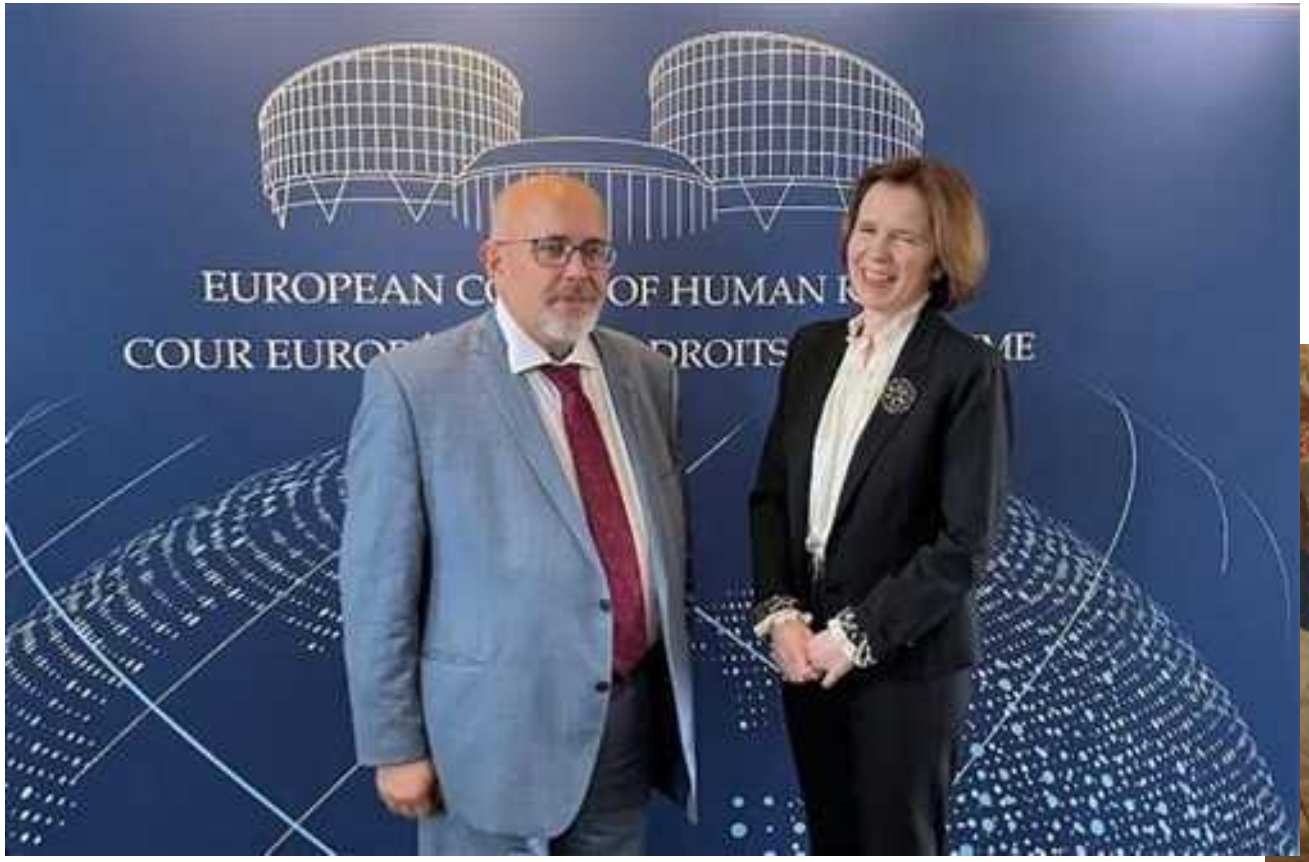
European Data Protection Supervisor (EDPS)

- **Privacy and data protection are fundamental rights** – see Articles 7 and 8 of the Charter of Fundamental Rights;
- **Independent supervision** is an integral part of the right to data protection – see Article 16(2) TFEU and 8(3) Charter;

What we do:

– monitoring and verifying compliance with Regulation (EU) 2018/1725,
– giving advice to controllers,
– handling complaints, conducting inspections,
– advising the co-legislators on new legislation,
– providing EDPB with secretariat,
– cooperating with Member States' DPAs,
– monitoring technological developments
– promoting data protection by design and development

European Data Protection Supervisor (EDPS)



European Data Protection Supervisor (EDPS)

The EDPS is an independent supervisory authority devoted to protecting personal data and privacy as well as promoting good practice in the EU institutions and bodies. A number of specific duties of the EDPS are laid down in Regulation 2018/1725.

Supervisory tasks **Cooperative tasks** **AI Market Surveillance Authority for EU institutions**

Consultative tasks **Secretariat of the European Data Protection Board.**

European Data Protection Supervisor 2004-2024

Two decades of personal data protection. What's next?

The EDPS' 20 Year Anniversary Book aims to chronicle how the EDPS emerged and developed as an EU body. It illustrates important milestones in the development of EU data protection law as a means to present the EDPS' contributions and activities. Taking stock of this rich history, the book includes contributions from key experts shedding light on the possible future of data protection and the challenges that lie ahead. Stay tuned for its publication.

https://www.edps.europa.eu/system/files/2024-06/edps_20thanniversary-book_en.pdf

20 talks for EDPS 20th annivesary

2 #20Talks - Nataša Pirc Musar

European Data Protection Supervisor • 256

3 #20Talks - Daniel J. Solove

European Data Protection Supervisor • 422

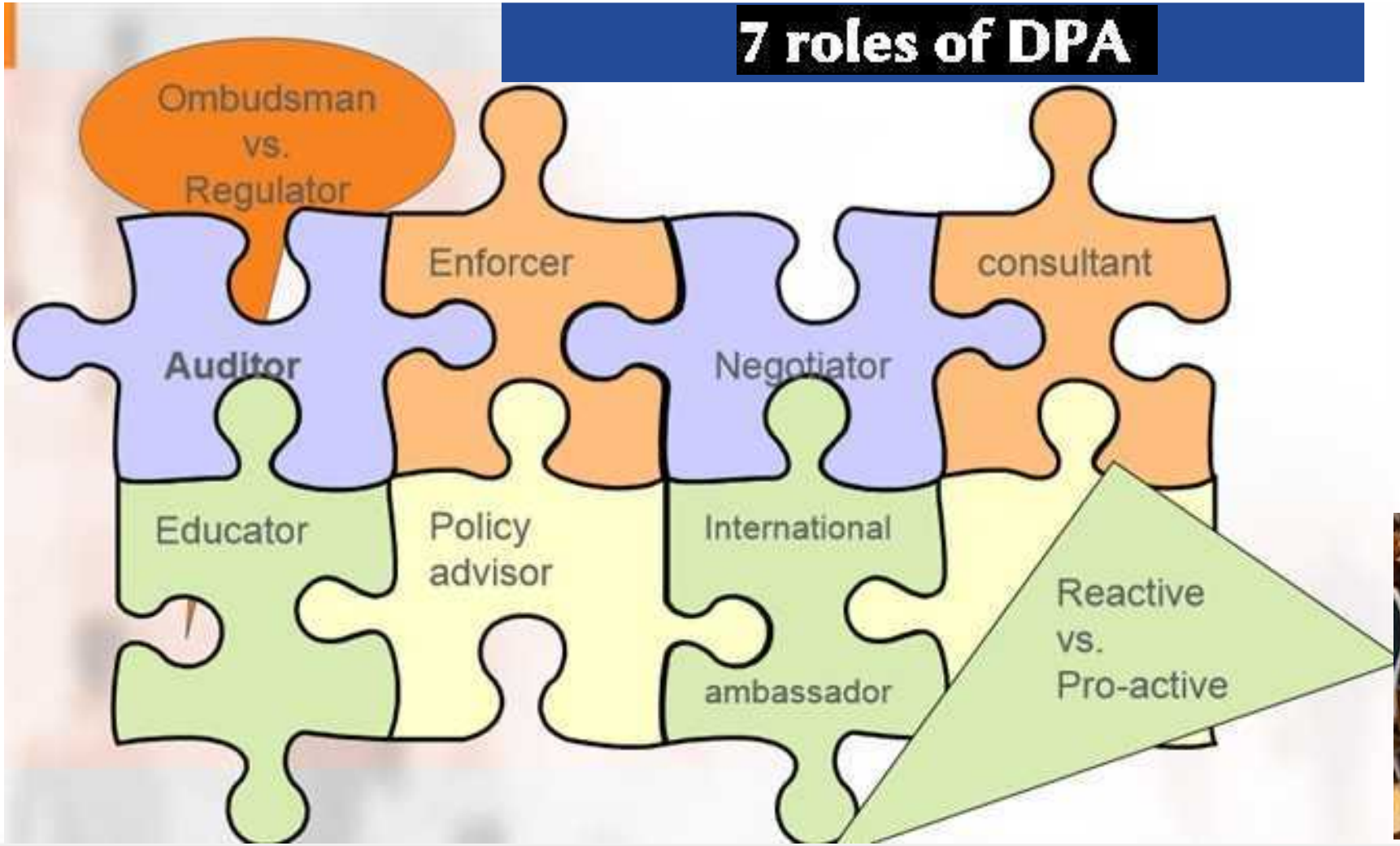
4 #20Talks - Koen Lenaerts

European Data Protection Supervisor • 249

5 #20Talks - Towela Nyirenda Jere

European Data Protection Supervisor • 55 vi

Data Protection Authorities



Colin Bennett, Charles D. Raab, *Revisiting 'The Governance of Privacy': Contemporary Policy Instruments in Global Perspective, 2018.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972086

The European Data Protection BOARD

- The European Data Protection Board (EDPB) is an independent European body, which contributes to the consistent application of data protection rules throughout EU, and promotes cooperation between the EU's data protection authorities. The EDPB is established by GDPR.
- The EDPB is composed of representatives of the EU national data protection authorities and EDPS).

- The supervisory authorities of the EFTA EEA States (IS, LI, NO) are also members with regard to GDPR-related matters and without the rights to vote EDPB has a Secretariat, which is provided by the EDPS. A Memorandum of Understanding determines the terms of cooperation EDPB and EDPS.

European fundamental right

Treaty on Functioning of European Union – Article 16

- 1. Everyone has the right to the protection of personal data concerning them.

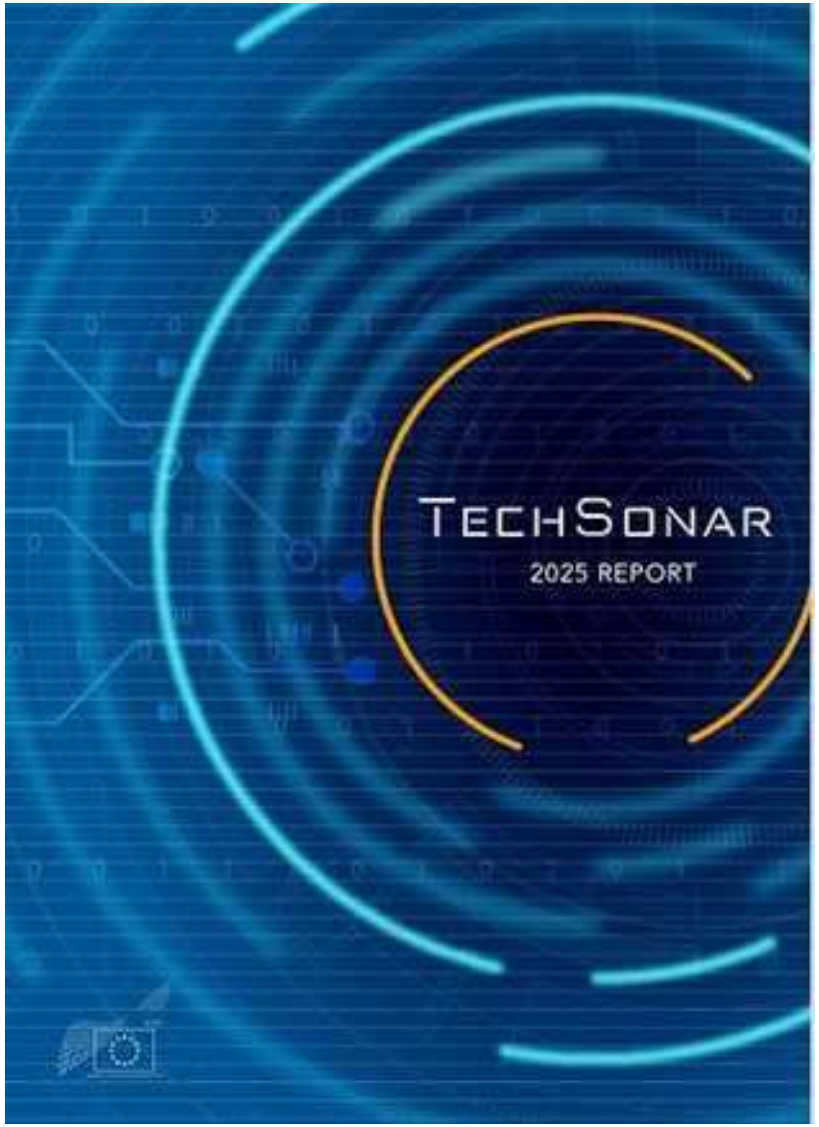
- 2. The European Parliament and the Council, acting in accordance with the ordinary legislative procedure, shall lay down the rules relating to the protection of individuals with regard to the processing of personal data by Union institutions, bodies, offices and agencies, and by the Member States when carrying out activities which fall within the scope of Union law, and the rules relating to the free movement of such data. Compliance with these rules shall be subject to the control of independent auhorities.

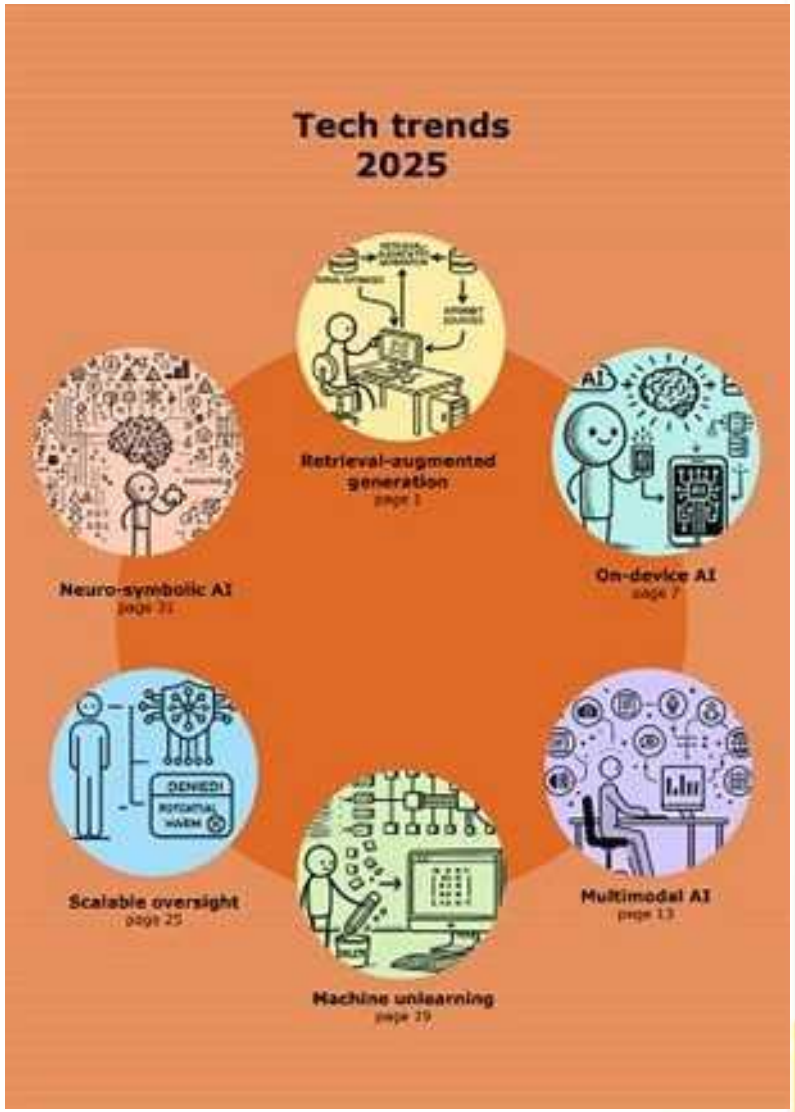
- 3. The rules adopted on the basis of this Article shall be without prejudice to the specific rules laid down in Article 39 of the Treaty on European Union.

Tech Sonar

The European Data Protection Supervisor (EDPS) regularly publishes TechSonar reports on emerging technologies.

TechSonar aims to anticipate emerging technology trends: the main aim of this initiative is to better understand future developments in the technology sector from a data protection perspective. Based on the collective intelligence of the EDPS staff, we aim to contribute to the wider debate on foresight within the European institutions.



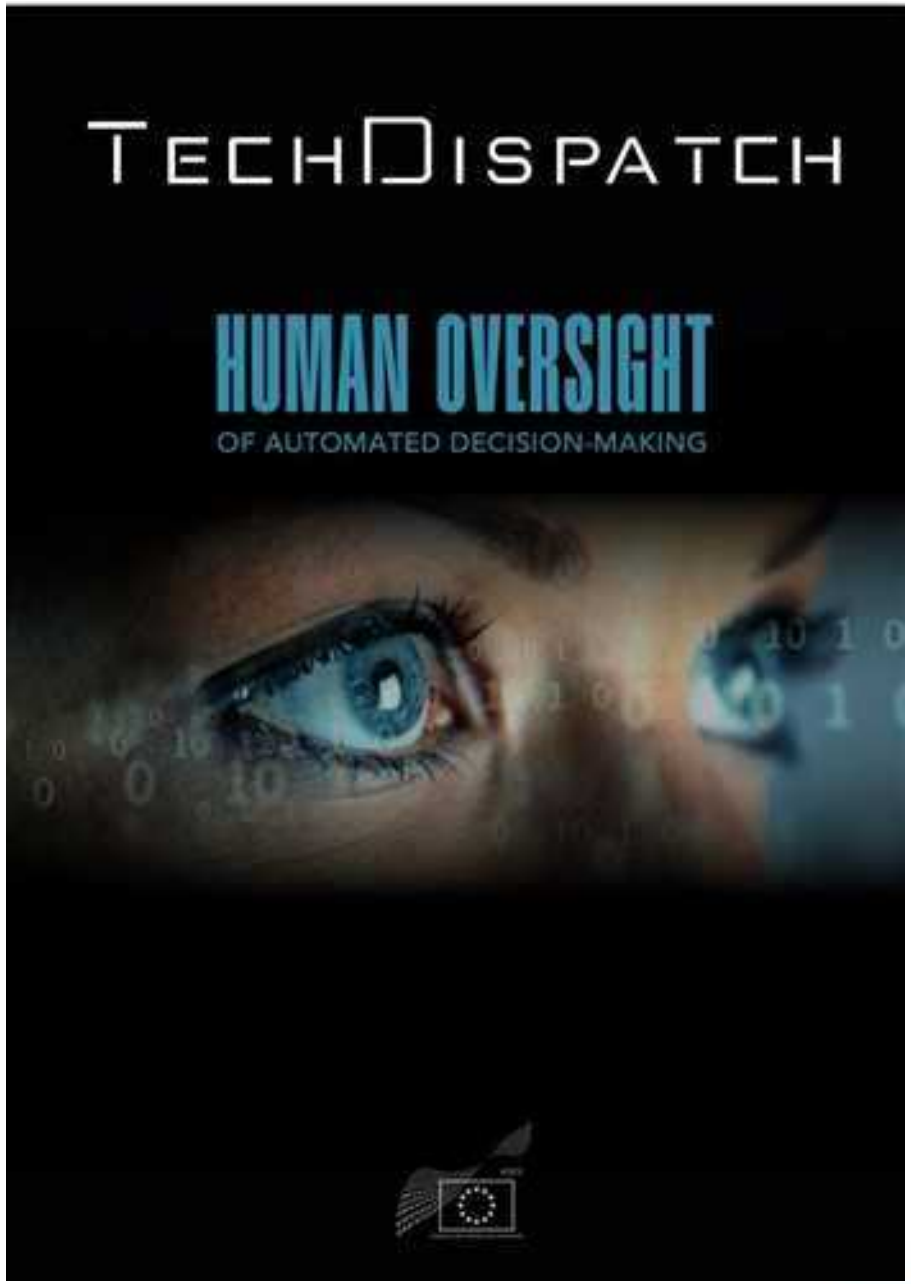


https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar_en

Tech Dispatch

The European Data Protection Supervisor (EDPS) regularly publishes TechDispatch reports that aim to explain emerging developments in technology. The TechDispatch reports are part of the wider EDPS activities on technology monitoring.

Each TechDispatch provides factual descriptions of a new technology, preliminarily assesses possible impacts on privacy and the protection of personal data, as we understand them now, and provides links to further recommended reading. Now you can also listen a podcast discussing main topics in 'TechDispatch Talks'.



https://www.edps.europa.eu/system/files/2025-09/25-09-15_techdispatch-human-oversight_en.pdf

"Human oversight"

Human oversight refers to the active involvement of at least one human *operator* in monitoring ADM system operations, evaluating the system's decisions, and having the ability to intervene if necessary

- *Sterz, S., et al. (2024). On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives. arXiv (Cornell University).*

While human oversight can occur at different stages of a system's lifecycle, including before deployment (ex-ante), *real-time* oversight on system operations (i.e. the supervision of systems deployed in production environments) is considered the one that can be most consequential.

The human 'operator' is a person who is properly trained to be able to supervise the system and has enough authority to override the system's decisions.

- *Eaton, M. L., & Illes, J. (2007). Commercialising cognitive neurotechnology-the ethical terrain. Nature Biotechnology, 25(4), pp. 393-397.*
- *OECD Recommendation on Responsible Innovation in Neurotechnology - OECD [Accessed: 10 January 2024].*

"Human oversight" according to AI Act

Article 14 Human oversight

*1. **High-risk AI systems** shall be **designed and developed** in such a way, including with appropriate human-machine interface tools, that they **can be effectively overseen by natural persons** during the period in which they are in use.*

2. Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular where such risks persist despite the application of other requirements set out in this Section.

*3. The oversight measures shall be commensurate with the **risks, level of autonomy and context of use** of the high-risk AI system, and shall be ensured through either one or both of the following types of measures:*

*(a) measures identified and **built**, when technically feasible, **into** the high-risk AI system by the provider before it is placed on the market or put into service;*

*(b) measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are **appropriate to be implemented by the deployer**.*

"Human oversight" according to AI Act

Article 14 Human oversight

*4. For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that **natural persons to whom human oversight is assigned** are enabled, as appropriate and proportionate:*

*(a) to **properly understand the relevant capacities** and **limitations** of the high-risk AI system and **be able to duly monitor its operation**, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;*

*(b) to **remain aware of the possible tendency of automatically relying or over-relying on the output** produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;*

*(c) **to correctly interpret** the high-risk AI system's **output**, taking into account, for example, the interpretation tools and methods available;*

*(d) to **decide**, in any particular situation, **not to use** the high-risk AI system or to **otherwise disregard, override or reverse** the output of the high-risk AI system;*

*(e) to **intervene** in the operation of the high-risk AI system or **interrupt** the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state.*

"Human oversight" according to AI Act

Article 14 Human oversight

*5. For [**remote biometric identification systems**], the measures referred to in paragraph 3 of this Article shall be such as to ensure that, in addition, no action or decision is taken by the deployer on the basis of the identification resulting from the system **unless that identification has been separately verified and confirmed by at least two natural persons with the necessary competence, training and authority**.*

*The requirement for a separate verification by at least two natural persons shall not apply to high-risk AI systems used for the purposes of law enforcement, migration, border control or asylum, where Union or national law considers the application of this requirement to be disproportionate **[???!!!]**.*

Automated decision making

However, while many examples focus on AI-based ADM systems, others do not.

It is important to emphasize that ADM systems can also be implemented without AI, using predefined rules, deterministic algorithms, or programmed logic.

We consider ADM systems in a broad sense - whether AI-based or not.

Article 22

Automated individual decision-making, including profiling

1. *The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*
2. *Paragraph 1 shall not apply if the decision:*
   - *(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;*
   - *(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or*
   - *(c) is based on the data subject's explicit consent.*
3. *In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.*
4. *Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.*

“Human oversight”

We will focus on ***real-time*** human oversight.

‘**Meaningful**’ human oversight is defined as active involvement that improves the quality of the decisions taken by the system, rather than as a merely procedural formality, or symbolic gesture. A Human is empowered and positioned to monitor the system in a substantive way.

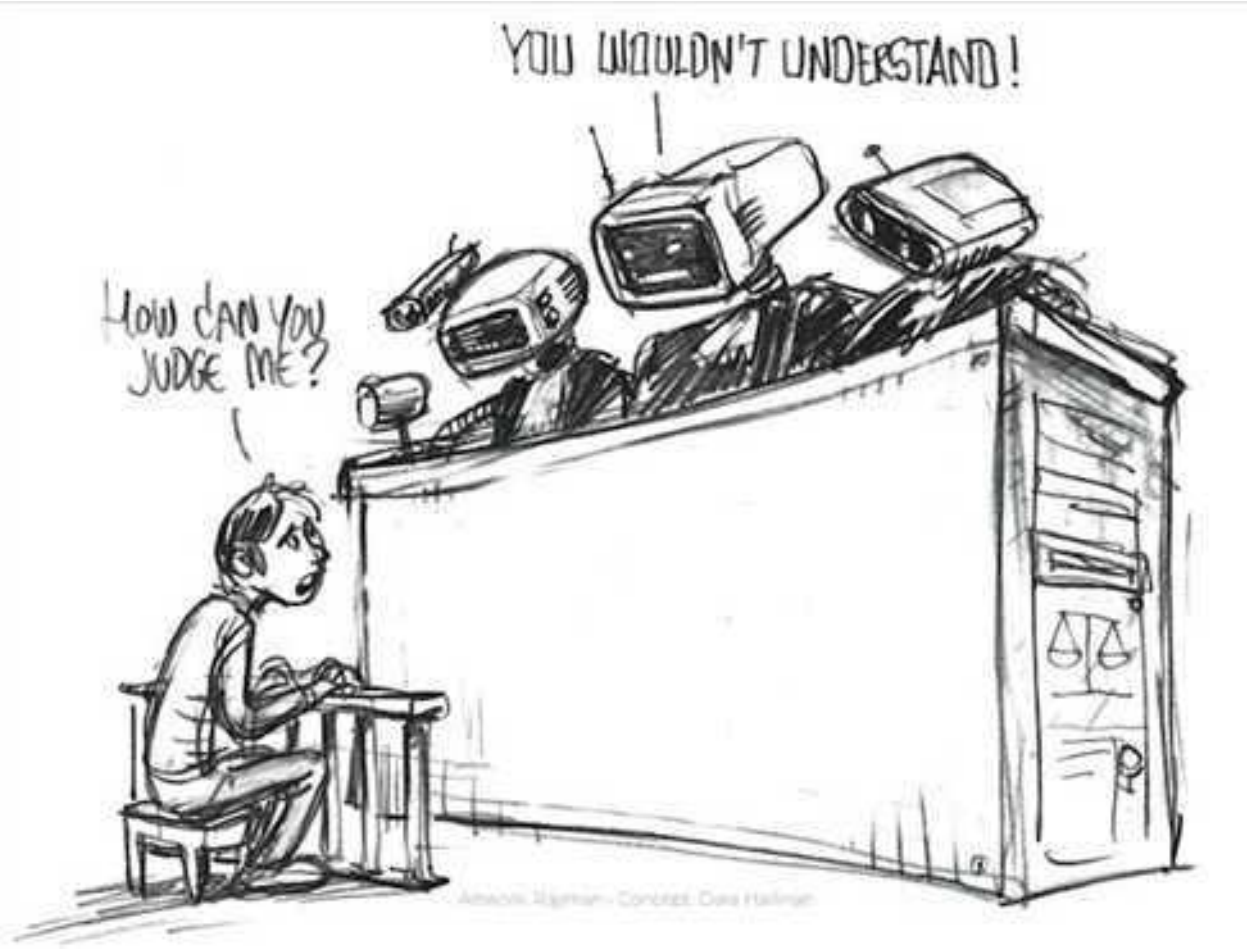
‘**Effective**’ human oversight refers to human involvement that has a tangible, positive impact on outcomes - specifically by preventing or mitigating harm, promoting fairness, and enhancing the overall quality, reliability, and accountability of decisions.

Examples of scenarios where human operators monitor systems designed to automate decision-making processes. The most prominent types of ADM systems are those powered by AI technologies - such as machine learning and deep learning - due to their advanced capabilities and widespread use across multiple sectors.


APPLICATION
DENIED

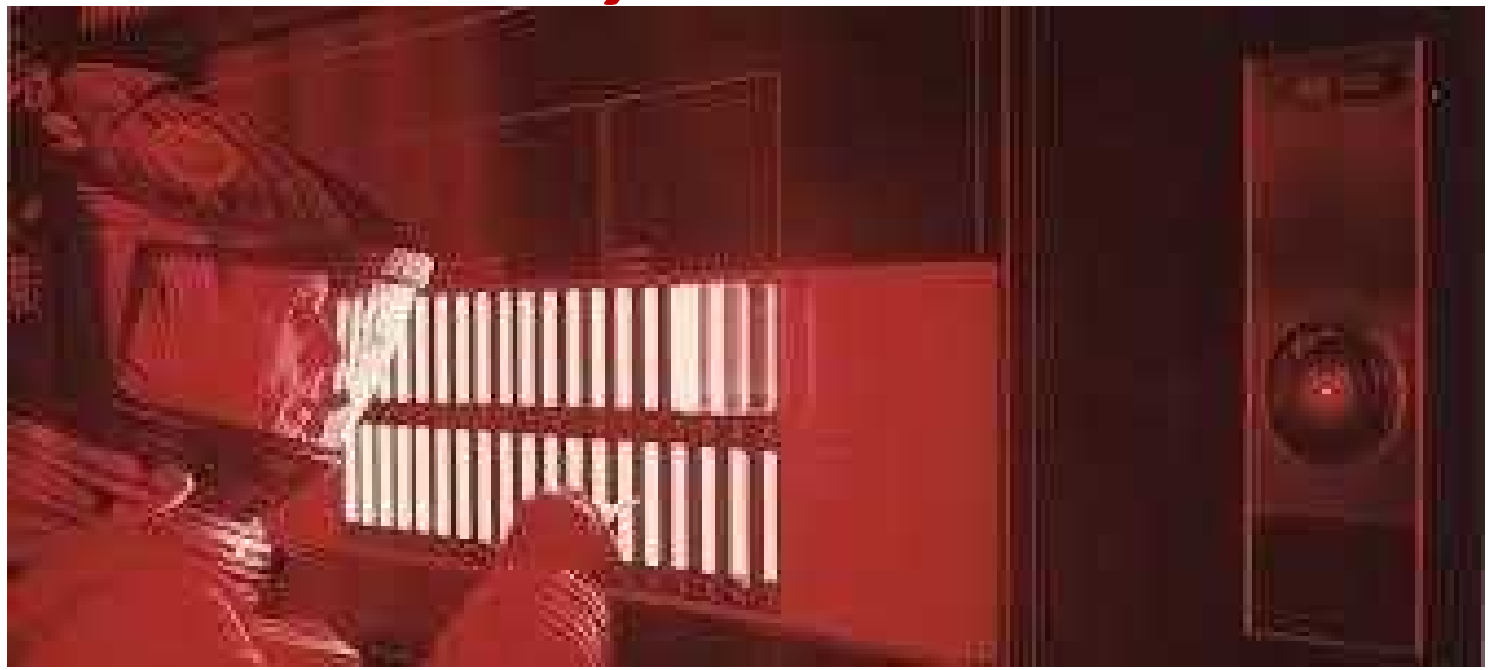
Human presence v. Human oversight

- We began by automating **muscle** through machines, engines, and robots.
- Then we automated **memory** through data, software, and the cloud.
- And now... we are automating **judgment** itself.



Human presence v. Human oversight

“There will always be a human in the loop.”



“I'm sorry Dave, I'm afraid I can't do that”

Wrong assumptions about the technology used in ADM systems

- ADM systems will operate within specific and predetermined conditions

- ADM systems will transfer control to humans when handling atypical or outlier situations

Wrong assumptions about the technology used in ADM systems

Uber self-driving car test driver pleads guilty to endangerment in pedestrian death case

Rafaela Vasquez, right, speaks with her attorney Albert Jaynes Morrison on Friday, July 28, 2023, in Phoenix.

Rafaela Vasquez was watching television on her smartphone in March 2018 when the Uber self-driving vehicle fatally struck Elaine Herzberg, 49, who was crossing a road in Tempe, Arizona, according to a National Transportation Safety Board investigation. Herzberg's death was the first known fatality involving a fully autonomous vehicle.

A judge in the Superior Court of Maricopa sentenced Rafaela Vasquez to three years of supervised probation and determined that the charge would only be designated a misdemeanor "upon successful completion of her sentence," according to a news release.

Wrong assumptions about how humans and technologies interact

- Automation does **not influence** human judgment
- There is no automated decision-making if there is a human supervising the system
- Human operators possess the **authority** to decide
- Systems **combining human and machine** work better
- Human operators have appropriate **mechanisms to override** the system's decisions
- Transparency and explainable AI improve human oversight
- Human oversight can help address the system's shortcomings

Wrong assumptions about how humans behave

- Human operators know what to do
- Human operators are able to do the task
- Human operators will have ‘fitting intentions’

Data protection at the foundation of trust

- The moral backbone of digital automation.
- Data protection principles
  - lawfulness, accuracy, purpose limitation, proportionality - pave the way for meaningful oversight.
- Know your data.
- Know why you use it.
- Know whether it still tells the truth.

Data protection at the foundation of trust

- This is precisely what the Council of Europe's Convention 108+ seeks to prevent –

Quiet erosion of fairness

through data misuse.

- As the world's only binding international treaty on data protection, it reminds us that privacy, accuracy, and accountability are not technical settings:

they are human rights.

Data protection at the foundation of trust

Human reviewer should therefore ask not only, “Is this outcome correct?”

but also, “Was the data lawful, necessary, and fair?”

Fairness begins long before the decision appears on a screen

- in the invisible chain of data that leads to it.

And that chain must be anchored in governance.

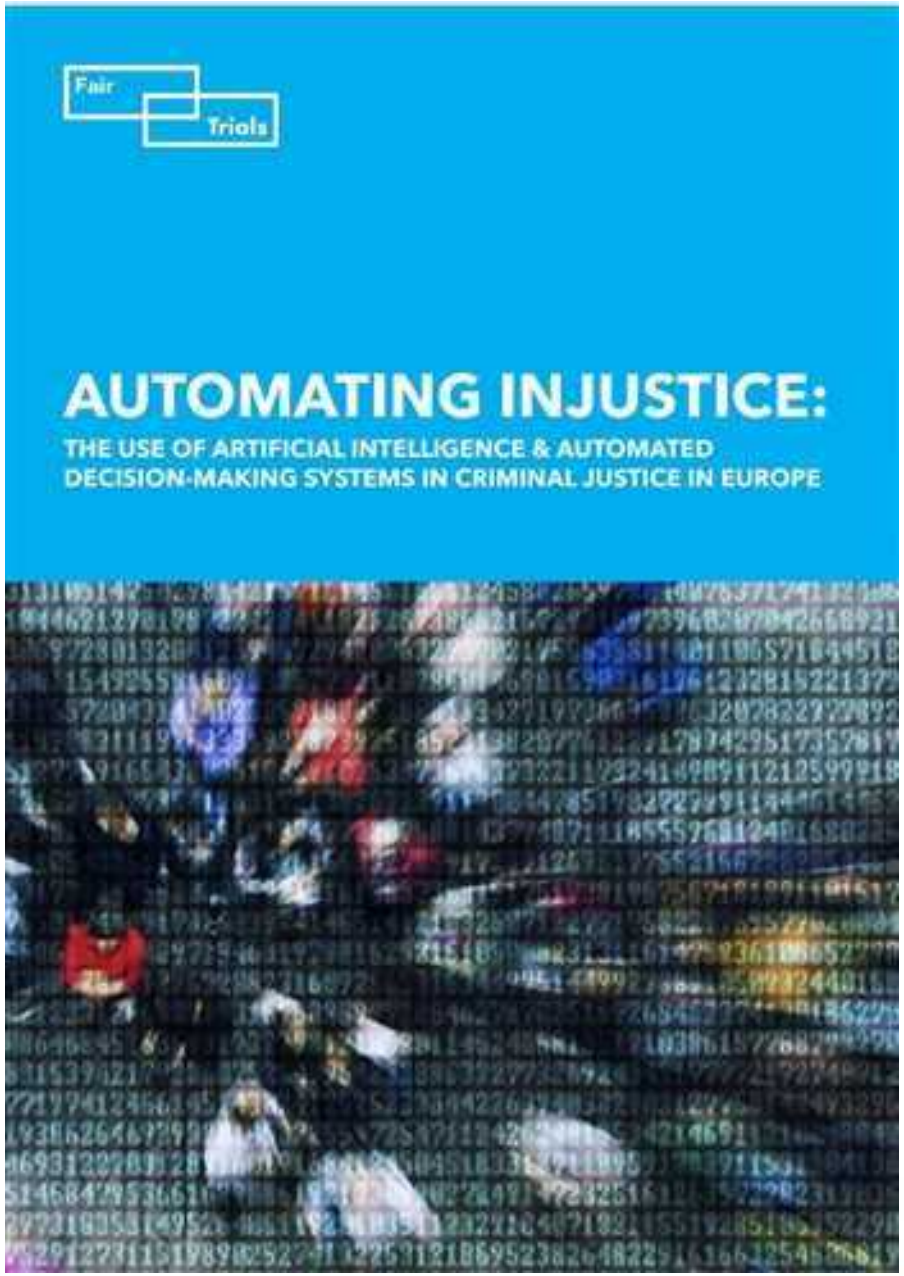
The paradox of oversight

	HUMANS	COMPAS
Accuracy (overall)	67.0%	65.2%
False positive (black defendants)	37.1%	40.4%
False positive (white defendants)	27.2%	25.4%
False negative (black defendants)	29.2%	30.9%
False negative (white defendants)	40.3%	47.9%

Humans score just as well as the **Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)** algorithm when they correctly predict someone's rearrest.

J. DRESSEL *ET AL.*, *SCIENCE ADVANCES*, EAAO5580, 2018, ADAPTED BY C. AYCOCK/*SCIENCE*

The paradox of oversight



The paradox of oversight

1.1 Human input and oversight: automation bias and legal loopholes

One of the main challenges of AI, automated, or semi-automated decision-making systems is that of 'automation bias' – the tendency to over-rely on automated processes in ways that can cause or errors in decision-making or legitimise errors or bias produced by an AI or ADM system.

Automation bias occurs primarily due to the perception that AI and ADM systems are generally neutral, reliable and, therefore, trustworthy. Automated evaluations have been found to be particularly salient to decision-makers, and research has shown that users of automated decision-making systems have a tendency to place greater weight on automated assessments over other sources of advice.

Most AI and ADM systems currently being used to inform or assist criminal justice decision-making do not completely replace human decision-making. They are instead designed and deployed to be used as decision aids, whose outputs are factored into consideration for the purposes of human decisionmaking. The phenomenon of automation bias however, raises questions about whether AI and ADM systems are being used in reality in accordance with their intended purpose as decision aids, and not as de facto replacements for human decision-making processes.

A simple requirement to have a human decisionmaker 'in the loop' or to have a human decisionmaker review or check the automated decision is insufficient, because this risks overestimating the capacity or willingness of human decision-makers to question and overrule automated decisions. A mere requirement to have an automated decision reviewed by a human, on its own, could reduce the human review into a rubber-stamping exercise which, in practice, is no oversight at all.

It is clear from the many examples considered above that AI and ADM systems currently being used in Europe often produce inaccurate, discriminatory outputs, despite the fact that there are human decisionmakers involved in these processes. This evidences how human intervention cannot compensate or correct for the problematic outcomes of an AI or ADM system.

Automating injustice: The use of artificial intelligence & automated decision-making systems in criminal justice in Europe; Fair Trials 2021, https://www.fairtrials.org/app/uploads/2021/11/Automating_Injustice.pdf

The paradox of oversight

- Presence is not power.
- Attention is not authority.
- A signature is not oversight.

The **“human in the loop” is either too slow, too uninformed, or too constrained to make a difference.**

What meaningful oversight really means

- True human oversight isn’t about inserting a person into a process

  - it’s about giving that person real agency to act on that system.

  Time to notice, **time** to think, **time** to act.

- Oversight only works when the human can step in before the damage is done.
  - How to “monitor” hundreds of algorithmic alerts per hour.

- When decisions happen in milliseconds, the right to intervene becomes theoretical.

- But time is useless without **understanding**.

What meaningful oversight really means

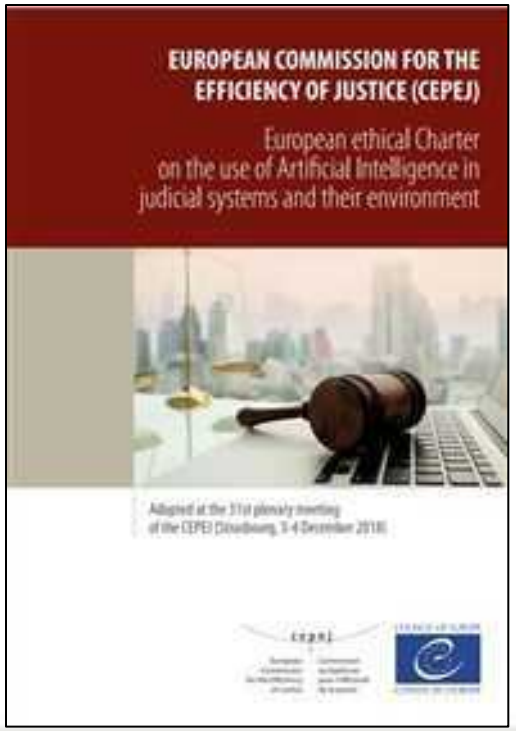
- **Agency** - the freedom to act on that understanding.
- A human overseer must be able to question or override a system when fairness or legality is in doubt.
- Oversight without agency is mere performance.

- Finally, oversight requires **intention**

- the moral compass that points the human toward justice rather than convenience.

CEPEJ Ethical Charter on the Use of Artificial Intelligence in Judicial Systems

The five principles of the Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment



1. **Principle of respect for fundamental rights:** ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights.

2. **Principle of non-discrimination:** specifically prevent the development or intensification of any discrimination between individuals or groups of individuals.

3. **Principle of quality and security:** with regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure technological environment.

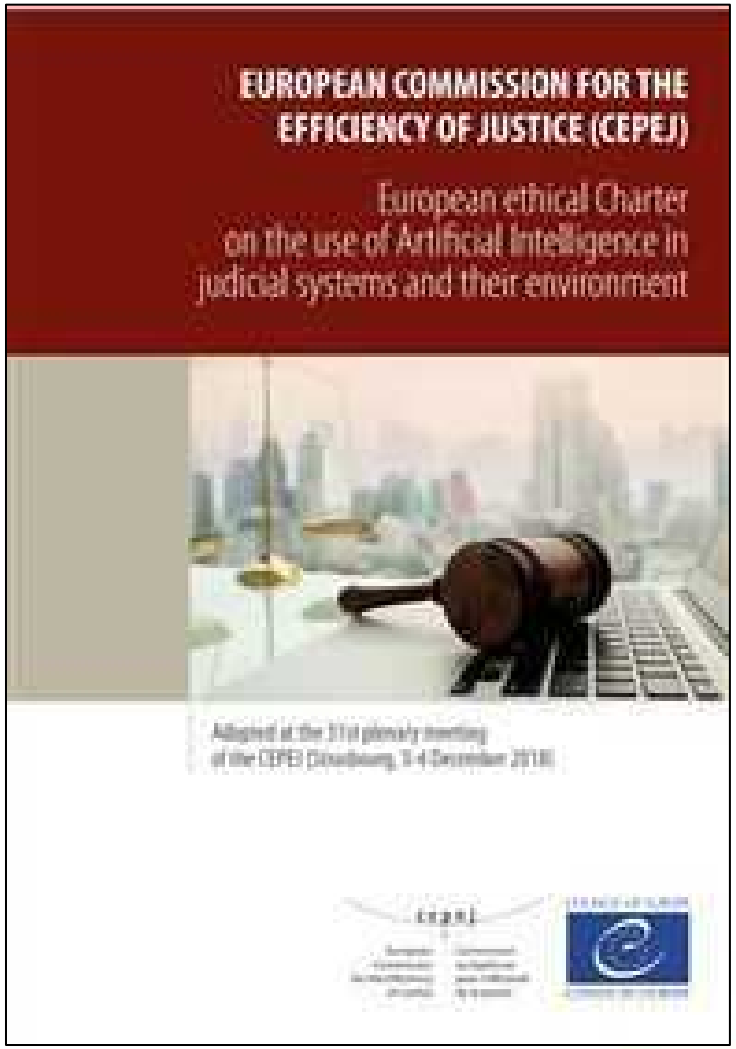
4. **Principle of transparency, impartiality and fairness:** make data processing methods accessible and understandable, authorise external audits.

5. **Principle "under user control":** preclude a prescriptive approach and ensure that users are informed actors and in control of the choices made.

CEPEJ Ethical Charter on the Use of Artificial Intelligence in Judicial Systems

- Five principles:
  - respect for fundamental rights,
  - non-discrimination,
  - quality and security,
  - transparency and impartiality, and
  - user control

- True oversight is not about catching errors after the fact. It's about shaping the conditions so that errors have less room to grow.
- It's about building systems that expect questions - and make space for them.



Can AIs explain judges' behaviour in retrospect?

An a posteriori explanation of a judge's behaviour, in particular the revelation of bias, would require all the potentially causative factorsto be identified through an interpretative framework and a contextualised analysis. The fact that, statistically, childcare is more often entrusted to mothers than fathers does not demonstrate a bias on the part of judges but reveals the need to mobilise different disciplines from the social scien[redacted]ht on this phenomenon.

Your text here

From a scientific point of view, explaining a phenomenon or, as far as we are concerned, a piece of human behaviour, amounts to determining the causal mechanisms that led to this behaviour using a certain amount of contextual data.

This requires, in a very schematic way, the preliminary constitution of an interpretative framework, itself derived from the repeated observation of this type of event or behaviour in the presence of certain factors or elements. The interpretative framework is made up of the hypotheses or points of view adopted by the different social science disciplines. This is an additional analytical step that can be fed into algorithms, but which they cannot perform alone.

X.Ronsin, V.Lampos, A.Maîtrepierre, In-depth study on the use of AI in judicial systems, notably AI applications processingjudicial decisions and data; CoE 2018; *https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment* , p.39-41.

When oversight cannot be continuous

Limit to how far human oversight can go.

- either automation grinds to a halt,
- or humans become mere rubber stamps, approving faster than they can think.

Neither outcome protects anyone.

Sampling?

- reviewing a representative set of decisions, regularly and rigorously.

Sampling allows us to detect drift, bias, or unintended harm before it spreads.

It turns oversight from an impossible full-time task into a sustainable, evidence-based practice.

When oversight cannot be continuous

Sampling alone is not enough.

Audit the auditors - both the system's actions and the human operator's actions.

- Did the algorithm behave consistently across groups?
- Did the humans intervene when they should have - or defer too quickly to automation?

Careful auditing creates the feedback loop that full-time human supervision can't provide.
It makes invisible processes visible again.

We cannot place a human in every loop, we must build loops that report back to humans.

Distributed oversight

- continuous monitoring by machines,
- reinforced by periodic, intentional review by people who understand both the data and the stakes.

This is how we preserve accountability at scale.

Not by pretending we can watch everything,

but by ensuring that nothing important goes unwatched

What to expect?

- The system should provide means for operators to intervene and, when necessary, override its decisions.
- The operator should have access to the relevant information needed to understand and evaluate the system's decisions.
- The operator should have agency and, as such, they must be able to override the decision of the system when appropriate (Green, 2022).
- The operator should have fitting intentions in line with their oversight role. For instance, operators should actively strive and mandate for the systems' decisions to be fair, unbiased and to respect fundamental rights.

Organisational measures

- Organisations should provide stable, healthy, and fair work conditions for adequate human oversight
- Organisations should provide appropriate and sufficient time for review
- Human oversight should be considered an important safeguard by the organisations
- Organisations should provide adequate training and support to the operators

Organisational measures

- The functioning of the ADM system. The operator should have sufficient understanding of the capabilities and limitations of the ADM system in place, as well as a general understanding of how the system works.

- A sample of cases of system failures, e.g., cases in which the recommendations provided by the ADM system were incorrect. The operators should be instructed on how to respond to - at least the most common - system failures. Research has shown that direct exposure to failures during training can reduce automation bias.

- The operational procedures (actions) that the operator is expected to perform within the system, such as approving or rejecting the output of the system. Operators might also need to provide feedback or an explanation if an output is rejected. Alternatively, the operators may be asked to provide a prior decision on the task before the system's automated decision is given, to better detect the presence of automation bias in tasks that are not time-sensitive.

- The tools that the operator can use to effectively perform its task. For example, the means that the operator has to regain control of the decision-making process or to correct the behaviour of the system.

Organisational measures

UNDERSTANDING AI USE BY COURT USERS

Courts will be in a better position to understand how use of AI may arise in court proceedings if they appreciate how different court users may approach the use of AI tools, as well as the associated level of risk for different uses.

- professionals must comply with certain **professional responsibilities**
- different levels of **technological and AI knowledge**
- a court user's **legal knowledge**

Technical measures for effective and meaningful human oversight

- The system decisions should be **explainable**
- The system **interface** should have a **clear, intuitive,** and **accessible design**

We will require... (1)

1. **Justification**. Organisations should be required to produce detailed reports justifying the deployment of ADM systems, including their necessity, goals, and expected outcomes.

2. **Periodic mandates**. Supervision roles should be subject to term limits or rotation (e.g., for auditors or review panels) to enhance impartiality and reduce the risk of alignment with provider or institutional interests.

3. **Collective decision-making.** Involving multiple individuals or bodies in supervision of decisions can reduce individual biases and disincentives. While it does not guarantee higher competence, it supports a more balanced and transparent evaluation process.

4. **Limited institutional competence.** Organisations should have narrowly defined human oversight mandates to prevent overreach or monopolisation of decisionmaking power.

5. **Justiciability and accountability.** Human oversight mechanisms must be subject to legal accountability. In liberal democracies, individuals must be able to challenge both public and private institutions in court to defend their rights, ensuring recourse in cases of algorithmic harm or injustice.

We will require... (2)

6. **Transparency**. Institutions must ensure meaningful algorithmic transparency, including:

- Clarifying whether human oversight is genuine or merely symbolic;
- Disclosing whether human oversight relies on AI and, if so, the data and methods used;
- Sharing the results of empirical testing on human oversight performance (as per the principle of justification);
- Describing the design and operational details of human oversight practices.

EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

@EU_EDPS

European Data Protection Supervisor

EDPS