Proving, not complying

Martijn van Grieken

September 2026

CONTENTS

Why the EU AI Act is an evidence regime, and how to get that evidence out of your systems. A framework for organisations that use AI in decisions about citizens, risks and patients.

Executive summary

On 27 July 2026 the Digital Omnibus on AI entered into force, six days before the date on which the AI Act's heaviest obligations were due to apply. The requirements for standalone high-risk AI systems now apply from 2 December 2027; for AI inside regulated products, from 2 August 2028. Many organisations have read that as breathing space. It is only half that. What has been postponed is the date on which a regulator can hold you to the high-risk requirements; the prohibitions, the transparency duties and the rules for general-purpose AI models already apply. What has not been postponed is your ability to answer it. And that ability cannot be arranged after the fact: a decision taken in 2026 can only be reconstructed in 2027 if it was recorded in 2026.

The heart of this whitepaper is a shift of perspective. The AI Act, and the frameworks that were already in place (the GDPR, Dutch administrative law, the Medical Devices Regulation, the case law on algorithmic decision-making), do not primarily ask you to comply. They ask you to be able to *show* that you comply, to a forum that will ask questions you do not yet know, about a moment that has passed. Complying is a state. Proving is a capability. And that capability is not a legal property of your organisation but a technical property of your systems: it lives in version control, logging, data lineage and the recording of human intervention.

We translate that into a framework. Every forum ultimately asks three questions: what was the system, what did it do, and who decided? And it asks them at two moments: beforehand, about the system as intended, and afterwards, about the system as it *actually was*. That gives an evidence matrix of six cells. Most compliance programmes fill the left-hand column: documents, assessments, declarations. Regulators, courts and appeals committees ask their questions in the right-hand column. Two mechanisms make that gap structural: evidence decay, because AI systems change without anyone deciding they should, and the oversight illusion, because human control exists on paper but in practice drifts along with the system. Both are measurable. We reduce them to five audit measures and close with five things you can do tomorrow, starting with one test: pick a decision from last quarter and try to reconstruct it. How long that takes is the most honest number you will get about your AI governance.

In public service delivery, security and regulated healthcare this is not a compliance matter at the margins. There, the answer to “how did this decision come about?” is a right held by the citizen, the suspect or the patient, and the duty to be able to give that answer existed well before the AI Act.

---

1. The deadline moved, the burden of proof did not

First the facts, because they shifted during 2026. The AI Act (Regulation (EU) 2024/1689) entered into force on 1 August 2024 and applies in phases. The obligations for high-risk AI systems were due to start on 2 August 2026. Because the harmonised standards and the national supervisory structures were not ready, the European Commission proposed the Digital Omnibus on AI in November 2025. After a political agreement in May and approval by Parliament and Council in June, it was published on 24 July 2026 as Regulation (EU) 2026/1744 and entered into force on 27 July.

The timeline, as at 1 September 2026:

Date	What applies
Since 2 February 2025	Prohibited AI practices (Art. 5). AI literacy (Art. 4); on 27 July 2026 the omnibus replaced that duty with an obligation of effort
Since 2 August 2025	Obligations for general-purpose AI models; governance and penalties
Since 2 August 2026	Transparency obligations (Art. 50): anyone interacting with an AI system must know it, and synthetic content, deepfakes and emotion recognition must be disclosed. For the marking of synthetic content, systems already on the market have until 2 December 2026
2 December 2026	New prohibitions on AI that generates non-consensual intimate imagery or child sexual abuse material
2 December 2027	Obligations for standalone high-risk AI systems (Annex III); previously 2 August 2026
2 August 2028	Obligations for high-risk AI in regulated products (Annex I), medical devices among them; previously 2 August 2027
2 August 2030	Final date for high-risk systems that were already running: see the transitional rule below (Art. 111)

Note the transitional rule for what is already running. A high-risk system placed on the market or put into service before the application date only falls under the requirements once it is substantially modified (Art. 111). That exception is not open-ended: providers and deployers of high-risk systems intended for use by public authorities must comply by 2 August 2030 at the latest. For public service delivery that is not an exemption but a longer run-up, and it is exactly the period in which the evidence has to come into being.

In the Netherlands the government put the AI Act Implementation Act out for consultation in April 2026. It places supervision with ten existing regulators, each in its own domain, with the Data Protection Authority and the Radiocommunications Agency as coordinators. That the national law arrives later than the European obligations changes nothing about those obligations: a regulation applies directly. And who the regulator turns out to be changes nothing about what you must be able to show.

Anyone who reads the postponement as nothing more than sixteen extra months misses three things.

First: the duty to prove was already there. Article 22 GDPR gives data subjects the right not to be subject to a solely automated decision producing legal or similarly significant effects, and where such a decision is permitted it requires suitable safeguards, the right to human intervention among them; Article 35 requires a data protection impact assessment. Dutch administrative law requires a decision to rest on sound and knowable reasoning. And the Dutch courts have already applied this to algorithms twice. In 2017 the Administrative Jurisdiction Division of the Council of State ruled, on the AERIUS calculation model, that a public body basing a decision on a computer model must disclose the choices made, the data used and the assumptions applied, in full, in time and on its own initiative, because otherwise the decision cannot be checked or challenged. In 2020 the District Court of The Hague declared the statutory basis of the SyRI fraud detection system inoperative for breaching Article 8 ECHR, partly because its workings were insufficiently transparent and verifiable (Wieringa, 2023). Neither ruling was about whether the state had done something wrong. Both were about whether that could be checked.

Second: evidence can only be built forwards. The AI Act requires deployers to keep the automatically generated logs of a high-risk system for at least six months (Art. 26(6)), and providers to keep technical documentation up to date (Art. 11). But the real reason to start now is not that retention period. It is that every decision taken today without a recorded data and model version will be permanently unreconstructable in December 2027. Postponing the enforcement date does not postpone the moment your evidence starts to exist.

Third: the backlog has been measured. In 2024 the Netherlands Court of Audit inventoried 433 AI systems across 70 central government organisations. For more than half, no risk assessment had been made; for 35 per cent of the systems in use, the organisation did not know whether the system was delivering what was expected; 5 per cent appeared in the public algorithm register. Two years earlier the Court had found risks in six of nine algorithms it examined, ranging from inadequate monitoring of performance to bias and unauthorised access (Netherlands Court of Audit, 2022, 2024). This is not a picture of organisations breaking the rules. It is a picture of organisations unable to show what they do.

2. Accountability is a relationship, not a document

The word “compliance” pushes thinking in the wrong direction. It suggests a state an organisation is either in or not, establishable with a checklist. What the law actually asks is better described in public administration scholarship. Accountability, in Mark Bovens’s definition, is a relationship between an actor and a forum, in which the actor is obliged to explain and justify their conduct, the forum can ask questions and pass judgement, and the actor may face consequences (Bovens, 2007). Three elements make that definition sharp for AI. The forum sets the questions, not the actor. The questions come afterwards. And the judgement follows not from what the actor intended but from what they can show.

For algorithmic systems there is a complication. From a systematic review of 242 publications on algorithmic accountability, Wieringa (2020) concludes that the “actor” behind an algorithm is rarely one person but a network of developers, suppliers, administrators and users, spread across time. Whoever is at the table when the question arrives usually did not build the system and often did not watch it change. Being accountable then means: being able to reconstruct, on behalf of that whole network, what happened.

This is also why transparency does not solve the problem. Ananny and Crawford (2018) show that seeing into a system, source code included, is something other than being able to interrogate it; being able to inspect a model tells you nothing about what it did with a given case file on a given day. Kroll and colleagues (2017) draw the consequence: accountability has to be built into a system’s design, with records that let you prove afterwards that a decision was reached by the procedure that was announced. The AI Act says essentially the same thing: a high-risk system must be technically designed so that events are automatically recorded over its lifetime, for the sake of traceability (Art. 12).

On our challenge page “We have to meet new requirements” we set out four questions an organisation can ask itself: do you know which data sits where, can you reconstruct a decision, is anyone accountable, and where does human intervention sit. Lay the questions of regulators, courts, auditors and appeals committees side by side and they reduce to three. **What was it?** Which system, which version, which data, which configuration. **What did it do?** Which output, on which input, and did it behave as validated. **Who decided?** Who was accountable, who reviewed it, with what mandate, and did that person depart from the output. And every forum asks those three at two moments: beforehand, about the system as intended, and afterwards, about the system as it was.

3. The evidence matrix: three questions, two moments

Cross the three questions with the two moments and six kinds of evidence appear. In brackets, the AI Act provisions the evidence falls back on; the list is illustrative rather than exhaustive; a few cells ask for more record-keeping than the provision literally requires, and which provisions apply to your system is for your lawyer to decide.

	Beforehand: the system as intended	**Afterwards: the system as it was**
What was it?	Technical documentation, data governance, description of training and test data, registration in the EU database (Art. 10, 11, 49)	Which version of model, data and configuration ran that day, and whether that state can be restored (Art. 12, 18)
What did it do?	Validation results, accuracy, robustness, known limitations and foreseeable misuse (Art. 9, 15)	Logs of input and output, drift monitoring, reported incidents (Art. 12, 19, 26(6), 72, 73)
Who decided?	Division of roles between provider and deployer, oversight assigned to people with competence and mandate, fundamental rights impact assessment (Art. 14, 26(2), 27)	Per decision: who reviewed it, what they saw, whether they departed from it and why; explanation to the data subject (Art. 26(11), 86; Art. 22 GDPR)

The left-hand column is where most compliance programmes put their energy. That is understandable: these are documents, they can be ticked off, they suit a project with an end date, and the conformity assessment explicitly asks for them. It is also the column the market of registers, assessments and certifications aims at. But a forum examining an individual decision almost always asks a question from the right-hand column. An appeals committee does not want to know what accuracy the model had at delivery, but what output it gave for this applicant, on the basis of which data, and whether the case handler still looked at it. A file describes the system as intended; a regulator asks about the system as it was, on that day, for that person, on that version.

The practical lesson of the matrix is as simple as it is uncomfortable: the left-hand column can be put right retrospectively, the right-hand column cannot. Missing documentation can be written. Missing logs cannot be generated. Whoever starts

filling the right-hand column only once enforcement begins will by then hold a file full of promises and no evidence at all.

From that follows the test we run first at every organisation, and that you can run yourself. Pick a decision from a random date last quarter in which an AI system played a part. With the people and systems you have today, try to establish which data went in, which version of the model was running, which output came out, and who then did what with it. Time it. We call this the reconstruction test. The result is usually not “impossible” but “three weeks and four people”, and that number is your starting position.

4. Two mechanisms every executive should know

Evidence decay: the system changes without anyone deciding it should. With a conventional application, a change in behaviour can almost always be traced to an intervention: a release, a configuration change, an updated dependency. With an AI system it does not. The data the system sees shifts: populations, forms, recording practices and policy change, and with them the distribution the model was once validated against. The literature separates a shift in the input distribution (dataset shift) from a shift in the relationship between input and output (concept drift); in practice the two blur together and the effect is the same (Gama, Žliobaitė, Bifet, Pechenizkiy, & Bouchachia, 2014) and has been described in clinical settings as the reason an approved model quietly degrades over time (Finlayson et al., 2021). Models are retrained, sometimes automatically. Upstream sources change their definitions without the consumer noticing; Sculley and colleagues (2015) called this the hidden technical debt of machine learning, summarised as: changing anything changes everything. And anyone building on an externally hosted language model receives new model versions from the supplier, without a release of their own and sometimes without notice. The conclusion is unforgiving: the distance between the documented and the running system grows by itself, and shrinks only through active intervention. Documentation is not a stock you build once but a flow you have to maintain. The regulation is explicit about this: risk management is a continuous, iterative process across the whole lifecycle (Art. 9), documentation must be kept current (Art. 11), and providers must set up post-market monitoring (Art. 72). A one-off assessment at go-live satisfies the letter and misses the point.

The oversight illusion: the human in the loop drifts along. Human oversight is the safety valve of almost every AI framework: Article 22 GDPR speaks of human intervention, Article 14 of the AI Act of human oversight, and nearly every

organisation has a case handler who “looks at it”. The problem is that people working daily with a well-performing system develop a tendency to follow it, including when it is wrong. That phenomenon, automation bias, has been documented experimentally since the late 1990s (Skitka, Mosier, & Burdick, 1999) and proves stubborn: it grows with workload and routine, and a warning alone does not remove it (Parasuraman & Manzey, 2010). Green (2022) examined policies requiring human oversight of government algorithms and concludes that they mainly work as legitimation: the system is deployed because a human looks at it, while that human is in no position to correct it.

The legislator knows this risk. Article 14 requires that people exercising oversight remain aware of the tendency to rely automatically on the system, can interpret outputs correctly, can decide not to use the system, and can intervene or stop it; Article 26 requires that this oversight be assigned to people with the competence, training and authority for it. Research into what makes oversight effective arrives at the same conditions: the overseer must genuinely be able to intervene, genuinely be able to see what the system is doing, and be able to govern themselves under pressure (Sterz et al., 2024); and oversight works only when it is designed from institutionalised distrust, with powers and counterweights, rather than as a courtesy (Laux, 2024). Looking on without mandate, time or information does not count. It is also rarely recorded anywhere, and that is the governance consequence: human oversight that is not logged does not exist as far as a forum is concerned. An oversight function that in ten thousand decisions never departs from the system does not prove the system is good; it is above all a reason to check whether anyone is actually overseeing. Take oversight seriously and you measure it.

5. From concern to baseline: five audit measures

Both mechanisms reduce to a small number of properties of a deployed system that you can measure, in an audit but also simply every quarter in operations. Not philosophy, but quantities.

1. **Reconstruction time.** How long it takes, for one decision on a random date, to establish the input data, the model and configuration version, the output and the human intervention. Target: hours, not weeks. This is the outcome of the reconstruction test, and the only measure that captures the whole matrix in a single number.

2. **Trail coverage.** The share of decisions for which a complete, immutable chain exists from input through version and output to human. A hundred per cent is the norm; every exception is a decision you cannot account for.
3. **Documentation lag.** The time between the last change to model, data or pipeline and the updated documentation. Zero is achievable when documentation is generated from the pipeline, as model cards and datasheets (Mitchell et al., 2019; Gebru et al., 2021), rather than written about it.
4. **Drift detection time.** The time between a shift arising in input distribution or performance and the moment anyone knows. If you do not know this measure, you find out about drift through complaints.
5. **Departure rate.** How often people exercising oversight depart from the system output, and how that is distributed across people, teams and periods. Zero is a red flag. A very high value is too: then the system is not trusted, or it is not usable. What is interesting is the range in between, and whether departures are reasoned.

These five measures are properties of your architecture, not of your legal documentation. They belong in the operational phase of every AI application, alongside the familiar monitoring of performance and security. And they are exactly the numbers a regulator, an auditor or a judge ultimately needs in order to get the three questions from section 2 answered.

6. What an organisation can do tomorrow

Five courses of action follow directly from the framework. They are written for organisations where AI systems contribute to decisions about citizens, risks and patients, and where the question “how did this come about?” is therefore not hypothetical.

1. **Complete the inventory and give every system a name.** Which AI systems exist, in which role (provider or deployer; substantially modifying a purchased system or putting it out under your own name can make you the provider), and in which risk class. The classification is work for your lawyer; the inventory is work for IT and the business together, and those two need to be looking at the same list. Attach one accountable person, by name, to each system. Shared accountability in practice means nobody has the overview at the moment it is needed.

2. **Run the reconstruction test, now.** One decision, one random date, one stopwatch. The result is your baseline on all five measures at once, and in our experience the most persuasive argument ever brought into a steering group on this subject. What now costs three weeks has to cost three hours by December 2027.
3. **Build traceability into the architecture, not into the file.** Version control of data, models, configurations and prompts; immutable logs with a retention period that covers at least the statutory six months and in the public sector usually has to run much longer, given appeal and archiving periods; data lineage carried through the whole chain; documentation generated from the pipeline. This is the same layer on which trust in your figures rests: definitions, lineage and ownership. Build that layer for accountability and you have built it for data quality too, and the other way round.
4. **Give oversight teeth, and measure it.** For each high-risk application, record who exercises oversight, with what mandate to depart, with what training, and with how much time per decision. Log every judgement by the person exercising oversight as an event in its own right, departures and their reasoning included. Design friction where it fits: interfaces that make the case handler form their own judgement before the system shows its suggestion demonstrably reduce blind acceptance (Buçinca, Malaya, & Gajos, 2021). And discuss the departure rate periodically, not as a performance measure for staff but as a health measure for the oversight itself.
5. **Treat evidence as a flow, not as a document.** Tie documentation to the release process, so that a change without updated documentation cannot be rolled out. Set up post-go-live monitoring for drift and incidents, with a reporting route that can serve the serious-incident obligation (Art. 73). Schedule revalidation on a fixed cadence and after every substantial change. And put all of this in operations, with a budget, rather than in a project that ends at go-live. The conformity assessment is where the evidence begins, not where it ends.

7. Three sectors, three emphases

Public service delivery. Here accountability is not a new obligation but the core of the craft: public bodies give reasons for decisions, citizens object, and the courts review. For public authorities the AI Act adds a fundamental rights impact assessment before a high-risk system is put into use (Art. 27), and, for most Annex III applications, the right of data subjects to an explanation of a decision with legal effect or similarly significant effect (Art. 86). The Netherlands moved early here with its algorithm register, the Fundamental Rights and Algorithms Impact Assessment

(Gerards, Schäfer, Vankan, & Muis, 2021) and the government's algorithm framework; the Court of Audit figures show that the instruments exist and the filling-in lags. The deeper point is older than any of these frameworks: Bovens and Zouridis (2002) already described how discretion in executive agencies shifts from the case handler to the system, and with it the place where accountability has to be organised. Fail to follow that shift and you get administration in which nobody can retell why a citizen received a decision.

Security. In the security domain a series of applications sits on the high-risk list, from risk assessment to evaluating the reliability of evidence, and predictive policing based solely on profiling is prohibited. At the same time, logs here are more sensitive than anywhere else, and separate regimes govern recording and disclosure. That does not make the right-hand column of the matrix less important but more decisive: where the use of a system cannot be public, reconstructability after the fact, for a court, a regulator or an internal review body, is the only thing left. Meijer and Wessels (2019), in their review of predictive policing, show that the promised effectiveness has rarely been demonstrated, while the risks to fundamental rights have been. Accountability here is also the way to prove that a system works.

Regulated healthcare. AI qualifying as a medical device already falls under the Medical Devices Regulation and becomes, under the AI Act, an Annex I high-risk system, with the new date of 2 August 2028 (Van Kolfschooten & Van Oirschot, 2024). For manufacturers that means one technical file for two regimes. For care institutions as deployers, the crux is that clinical validation is a snapshot, and that a model in practice is subject to dataset shift (Finlayson et al., 2021): different scanners, different populations, different recording. Healthcare already has a culture of evidence after go-live, through post-market surveillance and incident reporting. The question is whether that culture also covers the model versions, the input data and the clinician's judgement, or only the device.

8. What we do and do not cover, and an invitation

Twentynext does not give legal advice and does not determine which frameworks apply to your organisation, which risk class your systems fall into or which role you play in the chain. That belongs with your lawyer or compliance function. What we do is the technical side: making sure the questions they, and in due course the regulator, ask can actually be answered from your systems. The evidence matrix and the five audit measures are our instrument for that, and the reconstruction test is where we start.

Want to know where your organisation stands? We run the reconstruction test on one system of your choosing and turn the result into a baseline on the five measures, with an order of what has to come first. From a first audit through to setting it up and running it. Get in touch at info@twentynext.nl.

About the author

Martijn van Grieken is director of Data & AI at Twentynext, the Eindhoven-based data and AI consultancy that has worked for sectors including public service delivery, security and regulated healthcare since 2014. He has spent more than ten years at the intersection of data governance, AI and public decision-making.

About Twentynext

Twentynext is a data and AI consultancy based in Eindhoven, founded in 2014, supporting organisations in public service delivery, security and regulated healthcare, from seconding specialists to projects with a delivery commitment and ongoing operations.

Notes on sources

This whitepaper describes the state of the regulation as at 1 September 2026. References to articles of the AI Act concern Regulation (EU) 2024/1689 as amended by Regulation (EU) 2026/1744; the account is intended as orientation for decision-makers and is not legal advice. The evidence matrix, the five audit measures and the reconstruction test are instruments from our own practice; the underlying concepts (accountability as a relationship between actor and forum, automation bias, concept drift, hidden technical debt) come from the literature cited. All sources cited have been verified by the author.

AI language models were used in preparing this piece, under the editorial responsibility of the author.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review, 62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I. S., & Saria, S. (2021). The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine, 385*(3), 283–286. https://doi.org/10.1056/NEJMc2104626

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), Article 44. https://doi.org/10.1145/2523813

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Gerards, J., Schäfer, M. T., Vankan, A., & Muis, I. (2021). *Fundamental Rights and Algorithms Impact Assessment*. Utrecht University, commissioned by the Dutch Ministry of the Interior and Kingdom Relations.

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society, 39*(6), 2853–2866. https://doi.org/10.1007/s00146-023-01777-z

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration, 42*(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, 220–229. https://doi.org/10.1145/3287560.3287596

Netherlands Court of Audit. (2022). *Algoritmes getoetst: De inzet van 9 algoritmes bij de rijksoverheid* [An audit of 9 algorithms used by central government]. Algemene Rekenkamer. https://www.rekenkamer.nl/documenten/2022/05/18/algoritmes-getoetst

Netherlands Court of Audit. (2024). *Focus op AI bij de rijksoverheid* [Focus on AI in central government]. Algemene Rekenkamer. https://www.rekenkamer.nl/documenten/2024/10/16/focus-op-ai-bij-de-rijksoverheid

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems, 28*, 2503–2511.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., & Langer, M. (2024). On the quest for effectiveness in human oversight: Interdisciplinary perspectives. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 2495–2507. https://doi.org/10.1145/3630106.3659051

Van Kolfschooten, H., & Van Oirschot, J. (2024). The EU Artificial Intelligence Act (2024): Implications for healthcare. *Health Policy, 149*, 105152. https://doi.org/10.1016/j.healthpol.2024.105152

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*, 1–18. https://doi.org/10.1145/3351095.3372833

Wieringa, M. (2023). "Hey SyRI, tell me about algorithmic accountability": Lessons from a landmark case. *Data & Policy, 5*, e2. https://doi.org/10.1017/dap.2022.39

LEGISLATION AND CASE LAW

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *OJ L*, 2024/1689, 12.7.2024. http://data.europa.eu/eli/reg/2024/1689/oj

Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards simplifying the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). *OJ L*, 2026/1744, 24.7.2026. http://data.europa.eu/eli/reg/2026/1744/oj

Regulation (EU) 2016/679 (General Data Protection Regulation), in particular Articles 22 and 35.

Regulation (EU) 2017/745 on medical devices.

Administrative Jurisdiction Division of the Council of State, 17 May 2017, ECLI:NL:RVS:2017:1259 (AERIUS).

District Court of The Hague, 5 February 2020, ECLI:NL:RBDHA:2020:865 (SyRI).

Dutch Ministry of Economic Affairs. (2026). *Uitvoeringswet verordening artificiële intelligentie* [AI Act Implementation Act] (consultation version, April 2026). https://www.internetconsultatie.nl/uaiv/b1