Kleynen Consultants
SOFTWARE · GOVERNANCE · ACTUARIAL

MAY 2026

Transparency versus *trade secrets*.

The legal fault line in AI accountability for insurers.

Wouter Kleynen
AI Governance & Information Security

Adapted from Wouter Kleynen's talk at UNITAR's *"Everything AI"* event.
BRUSSELS · MAY 2026

unitar
United Nations Institute for Training and Research

INTRODUCTION

An AI system rejects an insurance claim. The policyholder asks a simple question: why? Answering it turns out to be hard, and not only for technical reasons. The question runs into a legal conflict between two legitimate interests. One is the interest in protecting how a system works. The other is the interest in understanding what it did. This article works through that conflict with a single insurance scenario set in the Dutch market. It shows where the two interests collide, why "just be more transparent" does not resolve it, and how EU law holds them in balance instead of sacrificing either.
The answer is not full public disclosure. It is layered access: enough for the policyholder to contest the decision, enough for the insurer to govern the system, enough for the supervisor to verify it, and enough protection to keep genuine trade secrets out of public view.

1 / THE FAULT LINE

Two interests, one *fault line*.

Trade secrets

The first interest is the protection of trade secrets. Under the EU Trade Secrets Directive, information qualifies as a trade secret only if it meets three conditions[1]:

- it is not generally known or readily accessible to people who normally deal with that kind of information;
- it has commercial value because it is secret;
- the person controlling it has taken reasonable steps to keep it secret.

Unlike a patent, which expires after twenty years and requires full public disclosure in exchange for protection, or a trademark, a trade secret has no registration requirement and no expiry date. It is protected for as long as secrecy is actively maintained[1, 19].

In insurance, two kinds of information are strong candidates for trade-secret protection, provided the three conditions above are actually met in practice. The first is fraud detection logic: the thresholds and rules a model uses to flag suspicious claims. The second is pricing models, where actuarial assumptions and parameters carry direct commercial value. Where the insurer or AI provider has taken real steps to keep these confidential, the company is legally entitled to protect both.

Transparency

The second interest is transparency. The EU AI Act describes transparency through four elements[2]:[A]

- **Traceability and explainability:** the system allows you to follow how an output was produced;
- **Awareness:** users know they are interacting with an AI system;
- **Deployer understanding:** the people running the system in production[B] know what it can and cannot do;
- **Affected-person information:** people on the receiving end know their rights.

The binding obligations in any given case come from the GDPR, DORA (the Digital Operational Resilience Act, a 2025 EU regulation that requires financial institutions to manage their IT third-party risks), Solvency II (the EU directive that sets governance and risk-management rules for insurers) and the AI Act itself. Different obligations attach to different categories of AI system.

> **A** I use the AI Act's transparency vocabulary throughout as an analytical lens, not as a single legal test that every AI system must satisfy in this exact form. The vocabulary captures four recurring elements of accountability; the binding force comes from the regulations.

> **B** In AI Act terminology, the “deployer” is the organisation using the AI system in its own operations, distinct from the “provider” that built and supplied it. In this article that is the insurer.

The legal fault line

These two interests pull in opposite directions. Trade-secret protection does not mean the information can never be disclosed. It means disclosure has to be controlled, legally justified and protected against unnecessary public or competitive exposure. Transparency means

disclosing enough of it that affected people, regulators and courts can understand what the system did and hold someone accountable when it goes wrong. Both interests are real, both are protected in law, and the two cannot be satisfied in full at the same time. That is the fault line, and the rest of this article is about how to stand on it.

2 / THE FRAME

Why maximum transparency is the *wrong frame*.

The intuitive answer, from the outside, is that more transparency is always better and trade secrets are the obstacle in the way of it. From inside an insurance company, that picture looks different: opening up everything about a system that took years to build also opens up a real loss. Both sides are correct about something, and that is why a blanket rule in either direction is wrong.

Trade secrets exist because full disclosure carries real cost. If a company publishes everything about a system it spent years building, two concrete things happen:

- **Its competitive position collapses.** Development choices that took years to accumulate become free for any competitor to copy.
- **The system becomes easier to attack.** Fraud-detection thresholds and security configurations are far harder to exploit when they are not public. Once they are published, fraudsters use them, and the resulting losses are spread across all policyholders as higher premiums.

But secrecy cuts the other way too. A trade secret can also conceal a system that is biased, unfair or unsafe. When "it is a trade secret" becomes a wall to dodge responsibility, that is the abuse the transparency rules exist to prevent.

> *The goal is balance. Organisations protect what really needs protecting, and affected people and supervisors get the information that really matters to them.*

3 / THE CASE

An insurer case, *worked through*.

AI use by insurers

De Nederlandsche Bank (DNB), the Dutch central bank and supervisor of insurers, published a sector-wide survey on insurance AI use in January 2026. Almost 80% of large and medium-sized Dutch insurers had at least one AI application in regular business processes at the start of 2025 (against around 21% of smaller insurers). DNB names two specific use areas: AI is being used for fraud detection and for assessing the size of claims, alongside broader efficiency and customer-experience applications. More than 70% of insurers say they take EIOPA's six AI governance principles into account, but the actual embedding of those principles is still developing[3].

The scenario in the next section is illustrative rather than a description of any specific deployment. The point is not what one insurer happened to do; it is to put the legal fault line under a single concrete fact pattern.

The scenario

A policyholder files a claim. On the face of it, the claim looks covered under their policy. A few days later, the insurer's response arrives: rejected.

The policyholder asks the insurer to explain the basis for the decision. The claim, it turns out, was processed with the help of an AI system. The underlying model was not built by the insurer; it was supplied by an external AI provider.[C]

ON LATER REVIEW

The rejection turns out to be wrong. The claim should have been approved. The AI produced an incorrect output, and the human reviewer did not catch it.

C I use “AI provider” throughout in the commercial sense: the external vendor supplying the system. Whether that vendor is also the “provider” of the deployed system under the AI Act,

> with the obligations that role carries, depends on a separate role analysis: who places the system on the market or puts it into service under their name, and whether the insurer substantially modifies or integrates it.

The insurer had not been careless. On paper, it had addressed each of the four transparency elements[2]:

- **Traceability and explainability.** The system logged the inputs it used and produced a structured reasoning record.
- **Awareness.** Policy documentation told policyholders that AI is used in claim handling.
- **Deployer understanding.** The claims team was trained on the system's intended use, accuracy range and limitations.
- **Affected-person information.** The rejection letter set out the right to request review and contest the outcome[10].

It had also built in a human check. A qualified employee reviewed every AI recommendation before any decision was sent out.[D] That step matters because under the GDPR, a person has the right not to be subject to a decision based solely on automated processing if the decision has a legal or similarly significant effect on them[4]. Refusing to pay out a covered claim is exactly that kind of effect.[E] A human review keeps the process out of the prohibition, but only if the review is real: the reviewer has to be able to read the case independently, understand why the AI recommended what it did, and have the authority to overturn it. A nominal sign-off does not count. If the human step is in name only, the decision is treated as solely automated and the GDPR safeguards apply[18]. In our scenario the employee did look at every case, which keeps the process technically on the right side of the rule; the problem, as we will see, is what *kind* of review that turned out to be.

> **D** The scenario assumes every claim is human-reviewed because that is where the legal question lands hardest. In practice, most insurers route only higher-impact or unusual claims to a full human review and let the simpler claims run through straight-through processing, with periodic sampling. Wherever a claim ends up being decided without meaningful human intervention, the GDPR analysis below applies.

> **E** "Decision" under Article 22 is read broadly enough to cover any automated output that effectively determines whether someone receives the benefit they contracted for, such as a credit score that decides whether a contract is signed or, in our scenario, an AI score that decides whether a claim is paid out[24].

Why the review failed

How does an error get through a process that was, on paper, complete and overseen? Two things went wrong at the same time.

- **Convincing but fabricated reasoning.** The system produced an explanation that read as coherent and matched the verdict it had given. On close reading, though, the explanation cited a policy clause that did not exist in the actual policy, and it described the contents of a receipt the policyholder had uploaded in a way that did not match what the receipt actually showed. The verdict was supported by reasoning that the model had effectively invented. This kind of failure is well documented across generative AI systems: they produce confident, fluent text that contains content the model has fabricated, a pattern often called *hallucination*[26]. From the outside the rejection looked properly justified. From inside the file, the justification did not hold up.

- **Automation bias.** A reviewer who sees the AI prove correct on most cases stops scrutinising it as hard. The literature on this effect goes back decades: reviewers tend to accept confident automated recommendations without checking them against the underlying facts, especially when the system has a strong track record[6]. The combination is the problem: when the explanation looks coherent and confident, and the reviewer is already inclined to trust the system, fabricated reasoning passes the eye test.

One named mitigation worth knowing about is *blind review*: the reviewer reaches their own decision on the case before seeing the AI's recommendation. It is not always feasible at volume, but it is the cleanest counter-design.

Every procedural step was completed, and the outcome was still wrong.

Why this is hard to investigate

The insurer can describe what happened: the model got it wrong, the reviewer missed it. For a supervisor that is the starting point, not the answer. The supervisor wants to know what structural features of the process made the failure possible, and what the insurer will change to prevent a repeat. Answering that means examining the model itself, and that is the hard part.

There is a real asymmetry underneath. When a human claims handler gets a case wrong, they can usually explain the error: a misread document, a wrong assumption, a missed exception. That explanation can be carried forward, so the same mistake is less likely next time. When the model gets a case wrong, neither step is reliable. The reasoning record the model produced may not faithfully reflect how the model actually arrived at its answer: research on chain-of-thought prompting shows that language models can produce explanations that read as coherent but do not reliably mirror what produced the answer[5], and in recent tests current

reasoning models acknowledged the factors that changed their answer in only around a quarter of cases[23]. Without a dependable explanation of *why* the system did what it did, there is no specific thing for the insurer to fix. The same input pattern can produce the same wrong answer next week, and nobody can point at the bug.

When the insurer asks the AI provider why

The insurer's natural next step is to ask the AI provider directly: on what basis did the model reach this conclusion?

The AI provider can usually supply useful evidence: input and output logs, model documentation, test results, performance by claim type, known limitations, feature-level indicators. What it often cannot supply, for complex machine-learning or large-language-model systems, is a complete account of why this exact output followed from this exact input. The provider's logs show *what* the model received and *what* it returned, not whether the explanation the model wrote afterwards is a faithful description of the actual computation. The governance question is not whether the provider can hand over every internal detail. It is whether the available evidence is enough for the insurer to judge whether the system is fit for this use.

Whatever the AI provider can or cannot show, the insurer's own accountability does not shrink because the system was built elsewhere. Three regimes make that explicit:

- **DNB supervisory expectations.** Since 2019, DNB has been clear that financial institutions remain fully accountable for the AI they use, whether built in-house or procured[7].
- **DORA.** Applicable since 17 January 2025, it requires the insurer to manage and oversee its IT third parties, including AI providers, and to hold contractual audit rights over them[8].
- **Solvency II.** It requires insurers to maintain sound governance and risk management over all systems material to their operations, including outsourced and third-party ones[9].

Telling the regulator that the AI provider would not explain is not, under any of these regimes, an answer the insurer can give.

Where trade secrets enter

So the insurer presses the AI provider for what it needs: the model's documented decision criteria, performance data by claim type, known limitations and failure patterns.

This is exactly the information that trade-secret protection is built to cover[1]. Disclosing decision criteria, training data composition or performance thresholds broadly would erode the AI provider's competitive position and, in fraud detection, its customers' security. The AI provider has real grounds to resist. The EU's highest court has recognised this directly: in a 2025 ruling, it held that a data subject's right to information about automated decision-making must be balanced against trade-secret protection, and that where the two conflict the contested information should go to the competent supervisory authority or a court rather than to the public[17].

Four parties now have a stake in the same claim: the policyholder, the insurer, the supervisor, and the AI provider. None holds the full answer alone. The court has already pointed at where the conflict is resolved: in a confidential supervisory or judicial channel, not in the open.

4 / THE STRUCTURE

Four parties, *one structure*.

The four parties want different things. The policyholder, the insurer and the supervisor each want information, in increasing depth. The AI provider wants protection. The structure has to deliver all four.

The policyholder

The policyholder needs process transparency. Depending on the facts, that may include confirmation that an automated system was used, what part it played, whether a human really reviewed the decision, meaningful information about the logic applied, the human-review route, and the right to contest the outcome[10]. The policyholder will not normally want or be entitled to raw thresholds, model weights or training data — those are technical details that do not help an individual contest a specific decision. That said, if the insurer or AI provider invokes trade-secret protection to resist disclosure, the contested material may need to be provided to the competent supervisory authority or court, which can then decide what, if anything, must be disclosed to the policyholder under controlled conditions[17].

The insurer

The insurer needs governance-grade understanding. To govern the system and meet its regulatory obligations, it needs enough understanding of the model to judge its fitness for the use case: how it was trained, its documented limitations, how its performance is monitored, and where its outputs need extra scrutiny. This is what the insurer must obtain from the provider under its DORA contractual rights[11].

The supervisor

The supervisor needs to verify the insurer's account independently, without taking it on trust. The route depends on the AI provider's legal role. If the AI provider is an ordinary IT third-party supplier, supervisory access usually runs through the insurer's own governance, records and contractual rights. If the AI provider is designated as a critical IT third-party provider under DORA, European supervisory authorities can act directly. If the AI provider acts as a processor or controller under the GDPR, the Dutch data protection authority (Autoriteit Persoonsgegevens, AP) can use GDPR investigative powers. If the AI provider is also the

"provider" of the system under the AI Act, AI Act market-surveillance powers apply once the relevant regime is in force[12]. Civil courts can also order limited disclosure to the opposing party under strict confidentiality, where necessary and proportionate[17].

The AI provider

The AI provider needs its secrets protected. Its interest is defensive: the decision criteria, training data and thresholds must not leak into open competitive view. The structure has to deliver the other three needs without forcing that.

> *An insurer that cannot explain its model to itself cannot explain it to a supervisor either. Transparency, here, means progressively deeper access.*

5 / THE RESOLUTION

How the *conflict is resolved*.

The policyholder

The policyholder who wants the individual decision corrected uses the insurer's complaints procedure first. If that does not solve it, the policyholder can take the case to Kifid, the Dutch ombudsman for consumer complaints against insurers and other financial-services providers, or to the civil courts[20]. Where the issue concerns personal data, profiling or automated decision-making, the policyholder can also complain to the AP, or seek compensation under the GDPR for material or non-material damage[4]. Under the AI Act, a person affected by a decision taken by a deployer of a high-risk AI system has a right to explanation of that decision once the high-risk regime takes effect, but only for systems that fall within the high-risk categories (for example, risk assessment or pricing in life and health insurance[15]) and not for every claim-handling support system[16].[F] For ordinary AI-assisted claim handling outside that perimeter, the binding routes are the GDPR, sectoral supervision, DORA, Solvency II, and the complaint and court channels described above.

> F The high-risk regime's original entry-into-application date was 2 August 2026. Following the May 2026 political agreement on the AI Omnibus, the planning baseline shifted to 2 December 2027 for stand-alone high-risk systems and 2 August 2028 for high-risk systems embedded in regulated products, pending formal adoption[15].

The insurer

The insurer's ability to re-examine a decision and withstand scrutiny rests on four things:

- **Contractual audit rights over the AI provider.** DORA's stronger audit regime applies if the AI service supports a critical or important function. Where it does, the contract must give the insurer real audit and inspection rights, exercisable by the insurer, an appointed third party or the competent authority[11]. Where it does not, the lighter general regime still requires written arrangements and third-party risk management. In practice direct site-level audits are usually impractical, and DORA itself permits reliance on pooled audits, third-party attestations and sector addenda where adequate[11]. But the entitlement still has to be real and enforceable on paper. The contract should also allocate liability, so the insurer has recourse if vendor failures lead to fines or customer harm.

- **Decision records under the insurer's assured control.** The insurer cannot depend on vendor-controlled logs alone. For API-served models, that means capturing the exact prompt or feature inputs, the context window, the model version, the output and the reviewer's action, rather than the model's internal computation, which the vendor will not expose. The insurer should also monitor for model drift: a vendor pushing silent weight updates can invalidate a previously sound governance position overnight, so periodic re-testing against a stable evaluation set belongs in the contract. This supports the record-keeping and governance obligations under Solvency II and DORA[8, 9]. Where the AI processing meets the threshold for high risk to data subjects, the insurer also has to carry out a Data Protection Impact Assessment under the GDPR and, if residual risk remains high, consult the AP before deploying the system[25].

- **A remediation and complaints route that works in practice.** The insurer must be able not only to reconstruct an individual decision but to correct it quickly, communicate the correction clearly, identify whether similar claims were affected, and decide whether the system should be paused, restricted or placed under enhanced human review. Bias and fairness monitoring belongs in the same loop: explainability is also the legal tool for demonstrating non-discrimination in pricing and triage, an obligation EIOPA expects insurers to operationalise across their AI portfolio[22].

- **The willingness to act on what re-examination finds**, including suspending or restricting AI-assisted decisions in any context where the system cannot be adequately explained.

In practice this requires a compact set of controls: an AI inventory; a role analysis under the AI Act, GDPR and DORA; classification of whether the service supports a critical or important function; a DPIA where the processing is likely to create high risk; contractual access and audit rights with liability allocation; insurer-controlled decision records; documented human review criteria and, where feasible, blind review for the highest-impact decisions; bias and fairness testing by claim type and protected characteristic; automation-bias training for reviewers; and a documented pause-and-escalate process.

If the insurer falls short, the consequences are legal as well as reputational, and the route depends on the failure:

- **AFM.** Where the failure points to conduct-of-business shortcomings or poor customer treatment[21].

- **DNB.** Where it points to weak governance, outsourcing, operational resilience or prudential risk management[8, 9].

- **The Autoriteit Persoonsgegevens.** Where personal data, profiling or automated decision-making is involved. The AP can impose fines of up to €20 million or 4% of total worldwide annual turnover[4, 13].

The supervisors

For supervisors, the trade-secret defence does not work as it does between commercial parties.

DNB supervises the insurer under Solvency II and DORA. It can examine the insurer's AI governance, review the contracts with the AI provider, and assess whether the contracted audit rights are adequate and actually exercised. For providers designated as Critical ICT Third-Party Providers under DORA, the relevant European Supervisory Authority can approach the provider directly: requesting information, conducting investigations, and carrying out on-site inspections. A non-compliant critical provider faces periodic penalty payments of up to 1% of its average daily worldwide turnover for each day of non-compliance, for up to six months[12].

The AP supervises the insurer as data controller under the GDPR. It can require the insurer, and any processor it uses (including the AI provider acting as a processor), to produce information about any processing involving personal data[13].

What reaches a supervisor does not become public. Professional-secrecy obligations bind supervisory authorities under both Solvency II and the GDPR[14]. Trade secrets can be shielded from broad commercial disclosure while remaining accessible to the supervisor through controlled channels, subject to professional secrecy and the limits of relevance and proportionality.

This layered approach is consistent with EIOPA's stance on insurance AI governance[22]. EIOPA's 2025 Opinion addresses insurance AI systems that are neither prohibited nor high-risk under the AI Act but still need governance under existing insurance-sector law: data governance, record-keeping, fairness, cybersecurity, explainability, human oversight and proportionality.

The AI provider

The provider's trade secrets survive this structure. They are protected from broad commercial disclosure: nothing here puts decision criteria, training data or thresholds into competitors' hands. What they are not protected from is regulated scrutiny: they can be required to pass into a confidential supervisory channel, bound by professional secrecy.

If the same provider is also the “provider” of a high-risk system under the AI Act (for example, one used for risk assessment and pricing in life or health insurance), AI Act supervisory powers also apply once the high-risk regime takes effect. National market surveillance authorities can require access to the technical documentation, training-data summaries and

conformity-assessment records, and non-compliance with the high-risk requirements carries fines of up to €15 million or 3% of worldwide annual turnover[15].

6 / CONCLUSION

Two legitimate interests, *balanced*.

The transparency / trade-secrets conflict is not solved by giving in on either side. The EU answer is structural. Four parties, each at a different layer of access. The policyholder gets enough to challenge the decision that affected them. The insurer gets enough to govern the system and answer for it. The supervisor gets enough to verify, under professional secrecy. The AI provider keeps what is genuinely commercial, while losing the option to hide behind it.

The work that remains is operational, not legislative. The legal scaffolding is in place. What it asks of insurers is concrete: an AI inventory, a real role analysis under the AI Act, GDPR and DORA, contracts that carry real audit rights and liability, decision records the insurer actually controls, human-review steps that are reviews and not sign-offs, and the discipline to pause a system when no one can explain what it just did.

For systems that fall within the AI Act's high-risk regime, the Act pushes that discipline into the design of the system rather than leaving it as an after-the-fact governance task[2]. For everything else, the same principles apply, just routed through GDPR, DORA and Solvency II.

> *The scenario in this article is a single wrong decision about a single claim. Behind it sits every structural question the sector will face at scale: who is accountable when the model is wrong, what the insurer actually controls, and whether the human in the loop is reviewing the case or ratifying the machine. The legal framework does not answer those questions for the insurer. It sets the conditions under which the insurer has to answer them itself.*

ABOUT THE AUTHOR

Wouter Kleynen

Wouter Kleynen is a Dutch specialist in AI governance with a background in mathematics and software engineering. In May 2026, he participated as a speaker on AI governance during the Everything AI event organized by UNITAR in Brussels. He is active in the fields of AI governance and information security at Kleynen Consultants. Since 1998, the company has specialized in developing software solutions, governance frameworks, and actuarial services, primarily for insurers and pension funds.

REFERENCES

Sources and *legal basis*.

[1] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets), OJ L 157/1, Article 2(1) (definition) and Article 5 (exceptions, including for legitimate interests recognised by Union or national law).

[2] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (the EU AI Act), OJ L 2024/1689, Recital 27 (citing the AI HLEG ethics guidelines; transparency defined as traceability and explainability, awareness of AI interaction, informing deployers of capabilities and limitations, and informing affected persons of their rights).

[3] De Nederlandsche Bank, "Verzekeraars en AI: DNB deelt nieuwste inzichten," Nieuwsbericht toezicht, published 21 January 2026.

[4] Regulation (EU) 2016/679 (General Data Protection Regulation), OJ L 119/1, Article 22(1) (the right not to be subject to a solely automated decision) and Article 82 (the right to compensation for material or non-material damage caused by an infringement).

[5] M. Turpin, J. Michael, E. Perez, S. R. Bowman, "Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting," *Advances in Neural Information Processing Systems* (NeurIPS) 36, 2023.

[6] L. J. Skitka, K. L. Mosier, M. Burdick, "Does automation bias decision-making?" *International Journal of Human-Computer Studies* 51, no. 5 (1999): 991–1006. See also K. Goddard, A. Roudsari, J. C. Wyatt, "Automation bias: a systematic review of frequency, effect mediators, and mitigators," *Journal of the American Medical Informatics Association* 19, no. 1 (2012): 121–127.

[7] De Nederlandsche Bank, "General principles for the use of Artificial Intelligence in the financial sector" (SAFEST framework: Soundness, Accountability, Fairness, Ethics, Skills, Transparency), 2019, principle on Accountability.

[8] Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (DORA), OJ L 333/1, Articles 28(1) and 30(2) (general contractual provisions for ICT third-party risk) and Article 64 (date of application: 17 January 2025).

[9] Directive 2009/138/EC (Solvency II), OJ L 335/1, Articles 41 (general governance requirements), 44 (risk management system) and 49 (outsourcing).

[10] GDPR Articles 13(2)(f), 14(2)(g) and 15(1)(h) (right to "meaningful information about the logic involved, as well as the significance and the envisaged consequences" of automated decision-making); GDPR Article 22(3) (safeguards including the right to obtain human intervention, to express one's point of view and to contest the decision).

[11] DORA Article 30(3)(e), requiring contractual provisions for "unrestricted rights of access, inspection and audit by the financial entity, or an appointed third party, and by the competent authority, and the right to take copies of relevant documentation on-site…" where the ICT services support a critical or important function. DORA Article 30(3) also permits reliance on third-party certifications and pooled audit reports where adequate. DORA Article 30(2) contains general contractual provisions applicable to all ICT third-party arrangements.

[12] DORA Articles 31–32 (designation of Critical ICT Third-Party Providers by the European Supervisory Authorities) and Articles 35–40 (powers of the Lead Overseer, including requests for information, general investigations and on-site inspections).

[13] GDPR Article 58(1) (investigative powers of supervisory authorities); GDPR Article 28 (processor obligations).

[14] Solvency II Directive Article 64 (professional secrecy obligations for supervisory authorities); GDPR Article 54(2) (professional secrecy obligations for supervisory authorities and their members).

[15] EU AI Act, Annex III, point 5(c) (AI systems intended to be used for risk assessment and pricing in relation to natural persons in the case of life and health insurance); Article 113 (the high-risk regime for Annex III systems applies from 2 August 2026); Article 99 (penalties).

[16] EU AI Act, Article 86 (right to explanation of individual decision-making by deployers of high-risk AI systems).

[17] Court of Justice of the European Union, Case C-203/22, *CK v Magistrat der Stadt Wien and Dun & Bradstreet Austria GmbH*, Judgment of the Court (First Chamber), 27 February 2025, ECLI:EU:C:2025:117, on the scope of GDPR Articles 15(1)(h) and 22 and the balancing of access rights with trade-secret protections.

[18] Article 29 Data Protection Working Party, "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679" (WP251 rev.01), adopted 3 October 2017 and last revised 6 February 2018, endorsed by the European Data Protection Board on 25 May 2018.

[19] Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Annex 1C to the Marrakesh Agreement Establishing the World Trade Organization, Article 39 (protection of undisclosed information).

[20] Klachteninstituut Financiële Dienstverlening (Kifid), "Een financiële klacht? Kifid kan in veel gevallen helpen," consumer guidance explaining that before a complaint is submitted to Kifid, it must first have been submitted in writing to the financial service provider.

[21] Autoriteit Financiële Markten (AFM), "Complaints," consumer guidance page, explaining the sequence: write to the financial undertaking, go to Kifid if unresolved, and consider court where appropriate; and noting that AFM does not mediate or decide whether individual complaints are well-founded.

[22] European Insurance and Occupational Pensions Authority (EIOPA), "Opinion on Artificial Intelligence Governance and Risk Management," EIOPA-BoS-25-360, 6 August 2025.

[23] "Reasoning Models Don't Always Say What They Think," Anthropic Alignment Science (preprint), arXiv:2505.05410, May 2025. Found that current reasoning models (including Claude 3.7 Sonnet and DeepSeek R1) acknowledged the factors that changed their answers in only ~20–25% of cases tested.

[24] Court of Justice of the European Union, Case C-634/21, *OQ v Land Hessen* (commonly cited as *SCHUFA Holding*), Judgment of the Court (First Chamber), 7 December 2023, ECLI:EU:C:2023:957, holding that an automated credit-scoring output qualifies as a "decision" under GDPR Article 22(1) where it effectively determines whether a third party concludes a contract with the data subject.

[25] GDPR Article 35 (Data Protection Impact Assessment for processing that is likely to result in a high risk to the rights and freedoms of natural persons) and Article 36 (prior consultation of the supervisory authority where residual risk after a DPIA remains high). See ref 4 for the consolidated text.

[26] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung, "Survey of Hallucination in Natural Language Generation," *ACM Computing Surveys* 55, no. 12 (2023): Article 248.