CHAPTER 6

Escalation Risk — four documented cases

> *Four public incidents in which unfilled roles did more damage than technical faults. In every one of them the machine executed the logic designed into it — the context around the machine was never finished.*

Each of the four cases realizes the same mechanism: an error or an unusual result on the model layer travels all the way to the layer of legal and financial consequences, because none of the layers in between — signal, operator, mandate, documentation, contract — stopped it on the way. The layers are not accidental; each case below shows exactly which one stood empty.

Once an AI incident has happened, a convenient wording lies close at hand: the model *hallucinated*, the model *broke*, the model *was not ready*. Convenient, because such wording moves responsibility onto the technology and away from the organizational decisions that preceded the incident. The more useful optic runs the other way: look at the decisions that were never taken and the roles that were never filled.

The four cases here are publicly documented incidents. Each carries a court judgment, a regulatory ruling or an official investigation with open materials. One thing unites them: in every case the technical part executed the logic designed into it — and the problem lived in the anthropological layer. In the question of who could and should have acted when the system produced a result that demanded a human reaction; in the architecture of roles, in the documentation, in the mandates, in the contracts. The technology never

slipped out of control — its context was never built to completion. An organization that assembles the technical layer without constructing the social one builds an escalation risk — the technology merely lends the unbuilt context speed and scale.

A different industry, a different scale, a different type of error in each case. And one identical pattern: a role unfilled, a mandate unclear, documentation missing or stale.

Norbert Wiener, in *Cybernetics*, formulated why this repeats systematically: an automatic system without a closed feedback loop to a human generates *drift* rather than errors in the narrow sense — a slow departure from intent that nobody registers until the critical failure.[1] Stuart Russell, in *Human Compatible*, translates this into the modern frame: escalation is a routine phase of the control loop rather than a crisis event, and it has to be designed — waiting for it to *fall into place by itself* is the design error.[2]

In each of the four cases below a specific role can be named that should have existed and did not — and precisely its absence turned the technically correct operation of a system into an incident. Case A demonstrates the missing *escalation role*: a person with an operational mandate to override the automation in a specific failure mode. Case B, the missing *editorial role*: a model curator who blocks the transfer of statistically predictive attributes from a commercial context into a state decision without separate justification. Case C, the missing *compliance translator*: a person who converts the model's internal decision into a form an external auditor can examine without breaching trade secrets. Case D, the missing *sovereignty auditor*: a role that tests how independent the contractual dependencies really are from geopolitical scenarios. Four industries, four error types — one shared structure of vacant roles. That structure, rather than the individual stories, is the subject.

Case A — aviation: the Boeing 737 MAX and MCAS

Context. The Boeing 737 MAX is an update of the base model in which a new system, MCAS — the Maneuvering Characteristics Augmentation System — was added to compensate for the aerodynamic changes brought by larger engines. In its original configuration MCAS took input from a single angle-of-attack sensor, which measures the angle between the wing and the oncoming air, and automatically pushed the aircraft's nose down to prevent a stall.

What happened. On 29 October 2018 Lion Air flight JT610 crashed into the Java Sea thirteen minutes after takeoff: 189 dead. A faulty angle-of-attack sensor fed an incorrect angle, and MCAS pushed the nose down in cycles. On 10 March 2019 Ethiopian Airlines flight ET302 crashed in the same scenario six minutes after takeoff: 157 dead. In total, 346 victims within five months. The worldwide grounding of the entire 737 MAX fleet lasted from March 2019 to November 2020.

What was missing. The operational mandate for manually overriding MCAS — when and how a pilot is obliged to switch the automation off, which signal counts as the trigger, how awareness is confirmed — was never written into the flight manuals. MCAS appeared in the documentation only minimally; training amounted to a tablet briefing of a few dozen minutes, with no simulator scenario of an active MCAS fed by a failed sensor. The cut-out switches physically existed, yet without prior exposure to this failure mode pilots had no trained response to it.

The anthropological reading. MCAS executed the logic designed into it — and that logic, the certification process, the design documentation and the pilot training proved insufficient for safe operation. After the grounding, the FAA approved a package of corrections: validation against two sensors, limits on repeated activation,

mandatory simulator training for every 737 MAX pilot, and an explicit manual-override protocol.[3] The US House Committee's final report of 15 September 2020 placed responsibility on the combination of Boeing's design decisions, certification shortcomings and the FAA's regulatory oversight.[4] The escalation role — the pilot as holder of the mandate for manual override in a specific failure mode — was secured neither by training nor by an explicit protocol; and that is only one of the layers of vacancy that accumulated toward the catastrophe.

Case B — public sector: the Austrian AMS algorithm, 2018–2020

Context. The Arbeitsmarktservice, AMS, is Austria's public employment service. In 2018 the AMS launched an algorithm meant to estimate, statistically, the chances of unemployed people returning to the labor market and to sort them into three groups: A with high chances, B with medium, C with low. The consultancy Synthesis Forschung built it; the project cost roughly 240,000 euros. The grouping was intended to inform decisions on access to retraining programs.

What happened. The first public audit of the model, published by AlgorithmWatch in 2019, showed that the formula assigned negative weights to the attributes *female*, *has children* — for women only, never for men — *age over 30*, *age over 50*, *disability* and *non-EU citizen*.[5] A woman with a child automatically scored lower than a man with an identical résumé. The AMS defended this as “a mirror of real chances on the labor market”. On 20 August 2020 the Austrian data protection authority, the Datenschutzbehörde or DSB, ruled in decision D213.1020/2020-0513.605 that the system was incompati-

ble with data protection requirements: no legal basis, no safeguards against discrimination, and a decision that was de facto automated despite being declared “support only”.[6] Deployment was suspended for the duration of the appeal. In 2025 Austria's Federal Administrative Court, the Bundesverwaltungsgericht or BVwG, set the DSB ruling aside — the court concluded that the AMS's data processing in this format does not violate the GDPR and that assessing discriminatory effect lies outside the DSB's competence.[7] The question remains open in the public and academic debate; legally, the system was never conclusively found unlawful.

What was missing. Before launch, nobody asked the question: *which correlations from historical labor-market data do we refuse to carry into a state decision about access to programs?* The team treated the task as technical — predict future chances from past data — while it was administrative: distribute a limited resource fairly. The role of a model curator in a state context, with the mandate to block attributes that correlate with protected categories even when they are statistically predictive, had no institutional anchor.

The anthropological reading. The model was built to predict the probability of returning to the labor market from historical data, and it demonstrated predictive power for exactly that indicator. The editorial role over the attributes — with the mandate to say “this attribute is admissible as a predictive variable in a commercial context, and it needs separate justification as the basis for a state decision” — stood vacant. The project cost taxpayers roughly 240,000 euros, produced a public scandal, a 2020 DSB ruling of incompatibility with the GDPR, and that ruling's reversal by the BVwG in 2025. Whatever the final legal outcome, the case remains the textbook example of how an administrative task — distributing a limited resource — can be dressed up as a technical one — predicting chances — and how the anthropological layer gets skipped in the process.

Case C — credit scoring: Schufa before the Court of Justice of the EU, Case C-634/21

Context. Schufa Holding AG, based in Wiesbaden, is Germany's largest credit reference agency. Its automated score, the Schufa-Score, is used widely by banks, leasing firms and telecom companies for decisions on credit, rentals and mobile contracts. The algorithm behind the score has never been published: Schufa invokes trade secrecy.

What happened. A citizen in Hesse was refused credit on the basis of a low Schufa score. She requested an explanation of the weights and factors. Schufa disclosed only the final number and a general description of the method, declining to reveal internal parameters. The administrative court of Wiesbaden referred the question to the Court of Justice of the European Union — the CJEU, the EU's highest court for the interpretation of European law. In its judgment in Case C-634/21 of 7 December 2023, the Court held: where a third party relies to a significant degree on an automated score in deciding about a person, that scoring falls under Article 22 of the GDPR — and requires mandatory safeguards: human review, a meaningful explanation, a right to contest.[8] A bank's formal “we merely take it into account” does not move the decision out of Article 22 if, in practice, refusal almost always follows a low score. After the judgment, credit bureaus across the EU revisited their transparency practice and their contestation rights.

What was missing. The system classified according to the internal specifications its business model is built on. The point of contact between it and an external examiner — first the customer, then a national court, then the CJEU — was never designed. No instrument for

explaining a model decision to a non-technical auditor. No mechanism that could meaningfully explain a scoring result without breaching trade secrets.

The anthropological reading. The scoring operated within its designed business logic. The role of a compliance translator — a person who converts the model's internal decision into a form an external auditor can verify without access to the model — existed neither at Schufa nor at most comparable agencies. Without that translation, automated scoring on which third parties significantly rely lands in the trap of Article 22 — the trap is sprung by the missing bridge between the model and a human with a control mandate, and never by the mere fact of automation. This is the classic vacant role of the AI operator in its regulatory-compliance dimension.

Case D — the 2022 cloud sanctions: AWS, Microsoft and Google stop service

Context. European industrial and financial companies run part of their machine-learning pipeline at public cloud providers — AWS, Microsoft Azure, Google Cloud — for economies of scale and access to managed AI and machine-learning services. Many of them had subsidiaries in Russia and Belarus, or partners operating in sanctioned jurisdictions.

What happened. On 4 March 2022 Microsoft announced the suspension of new sales of products and services in Russia.[9] On 8 March 2022 AWS stated it was no longer registering new customers in Russia and Belarus.[10] Google stopped new subscriptions to paid Google Cloud services. In December 2023 the EU's twelfth sanctions package — Council Decision CFSP 2023/2874 and Council

Regulation EU 2023/2878 of 18 December 2023 — extended the restrictions to the supply of enterprise software to legal entities registered in Russia.[11] As a result, a number of European companies with infrastructure shared across Russian subsidiaries had to rework the architecture of their international information systems and their procedures for cloud access. Public technical post-mortems — published analyses of what broke and how it was rebuilt — barely exist for these migrations. That is exactly why the case matters as the story of a new class of risk that surfaced after the 2022 sanctions decisions rather than as the story of any single company.

What was missing. Until 2022, the standard contract review for cloud services covered service-level agreements — the guaranteed availability and response times of a service — plus price, technical compatibility and GDPR conformity. The scenario of a geopolitically motivated denial of service — arising from the provider's legal obligation to its own state rather than from a technical failure — was covered in most European companies by no playbook, no contingency plan and no exit strategy.

The anthropological reading. Technically, the systems worked correctly up to the moment of the contractual or legal restriction. The role of a sovereignty auditor — a person who tests how independent the contractual dependencies really are from geopolitical scenarios and builds a contingency plan for a sudden cut-off — was never assigned. Procurement selected by price and service-level agreement; IT architecture by technical fit; legal by standard clauses. None of these departments held the mandate to ask: *under which scenarios can this provider legally lose the right to serve us — and what then?*

The shared pattern: in all four cases the problem was anthropology, never the model

Set side by side, the four cases read as one table.

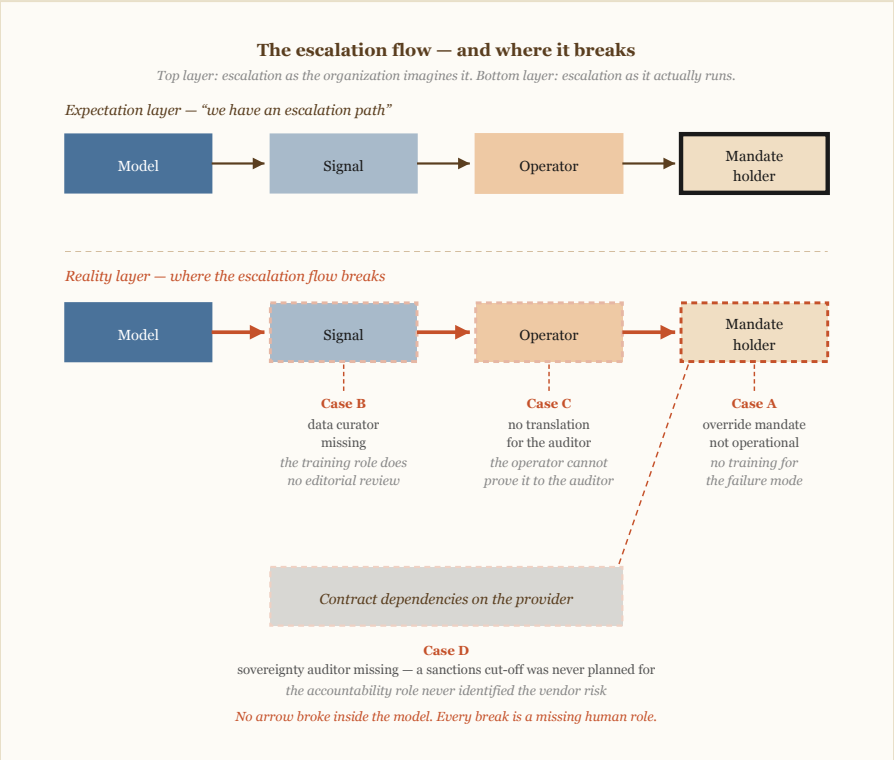
Case A — Boeing MCAS. The mandate for manually overriding the automation in a failure mode was not operational. Technically, MCAS executed the logic designed into it — while that logic, the certification and the documentation were drawn up without an explicit protocol for pilot action in the specific failure mode. The escalation role — the pilot as holder of the override mandate — was secured neither by training nor by protocol.

Case B — the AMS algorithm. A curator of attributes with the mandate to say "this variable is admissible as a predictive factor in commerce, and it needs separate justification in a state decision" was absent. Technically, the model performed the prediction it was built for. The editorial role over attributes in an administrative context stood vacant. The legal fate of the rulings was contested and reversed; the anthropological lesson was not.

Case C — Schufa before the CJEU. A mechanism of compliance translation for external auditors was absent. Technically, the scoring ran on its internal business logic. The bridge between the model and a human with a control mandate — customer, regulator, court — was missing, and exactly that turns automated scoring, under certain conditions, into an Article 22 case under the GDPR.

Case D — the 2022 cloud sanctions. A sovereignty audit of contractual dependencies was absent. Technically, the systems performed their designed function up to the legal restriction on the provider's side. The accountability role never identified the geopolitical risk inside the supplier dependencies.

> *In all four cases the model did, or tried to do, what it was designed for. The problem arose where the human structure — roles, mandates, documentation, contracts — gave nobody the context to react to what the system produced.*



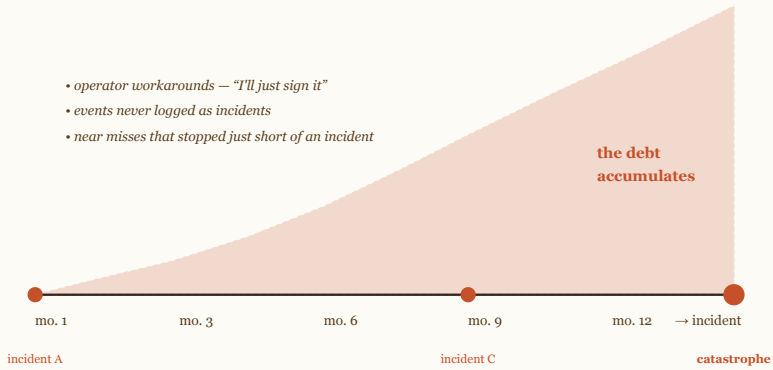
Four escalation cases A–D. Every break happened on a human layer — signal, operator, mandate, contract — while the model layer kept working as designed.

That is the point of the anthropological approach: it keeps the technical view and adds the questions technical analysis never asks.

Invisible escalation debt

the unseen debt of escalations that accumulates between incidents

"escalation debt" — the workarounds nobody wrote down



Visible incidents are points. The invisible debt is the area under the curve.

Between visible incidents the organization keeps accumulating unlogged workarounds and near misses — the debt that pays out all at once.

EXPECTED VERSUS REAL CONTROL

recurring motif · chapters 6 · 8 · 16 · 21

Expectation layer

human in the loop · an expert verifies · a manager signs off

the gap

Reality layer

a memo instead of an audit trail · a signature without review · a mandate without a lever

Control exists where a person has a mandate, a channel and time — a name in a document is only a name.

Three Questions for Your Own Context

1. **Mandates.** Does every escalation level of your system have named people in writing, with a clear decision mandate — including outside working hours?
2. **Data curation.** Who examines your training and retraining datasets for the anomalies the model *must not* learn?
3. **Supplier sovereignty.** Which of your AI components depend on a provider that, in a geopolitical scenario, can be legally forced to restrict access — and does your contingency plan cover that scenario?

Sources

1. Norbert Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 1948.
2. Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, 2019.
3. Federal Aviation Administration, *Summary of the FAA's Review of the Boeing 737 MAX*, 18 November 2020.
4. US House Committee on Transportation and Infrastructure, *Final Committee Report on the Design, Development, and Certification of the Boeing 737 MAX*, 15 September 2020.
5. Nicolas Kayser-Bril, “Austria's employment agency rolls out discriminatory algorithm, sees no problem”, AlgorithmWatch, 6 October 2019.
6. Datenschutzbehörde Österreich, decision D213.1020/2020-0513.605 of 20 August 2020.
7. Bundesverwaltungsgericht Österreich, ruling setting aside the DSB decision on the AMS algorithm, 2025.

8. Judgment of the Court (First Chamber) of 7 December 2023, Case C-634/21, *SCHUFA Holding (Scoring)*, ECLI:EU:C:2023:957.

9. Brad Smith, “Microsoft suspends new sales in Russia”, Microsoft On the Issues Blog, 4 March 2022.

10. Amazon Staff, “Amazon's assistance in Ukraine”, aboutamazon.com, 8 March 2022.

11. Council Decision (CFSP) 2023/2874 of 18 December 2023 amending Decision 2014/512/CFSP; Council Regulation (EU) 2023/2878 of 18 December 2023 amending Regulation (EU) No 833/2014 concerning restrictive measures in view of Russia's actions destabilising the situation in Ukraine, EUR-Lex.