

KAPITEL 6

Eskalationsrisiko — vier dokumentierte Fälle

Vier öffentlich dokumentierte Vorfälle, in denen unklare Rollen mehr Schaden angerichtet haben als technische Fehler.

Jeder der vier Fälle ist eine konkrete Ausprägung desselben Mechanismus: Ein Fehler oder ein unerwartetes Ergebnis auf der Modellebene erreicht die Ebene juristischer und finanzieller Folgen, weil keine der Zwischenschichten — Signal, Operator, Mandat, Dokumentation, Vertrag — die Ausbreitung gestoppt hat. Die Schichten sind nicht zufällig gewählt; jeder Fall zeigt, welche von ihnen leer geblieben ist.

Es gibt eine bequeme Erzählweise über KI-Vorfälle: Das Modell hat *halluziniert*, das Modell war *kaputt*, das Modell war *nicht bereit*. Die Erzählung ist bequem, weil sie die Verantwortung auf die Technologie schiebt — und sie von den organisatorischen Entscheidungen löst, die dem Vorfall vorausgingen. Dieses Kapitel dreht die Optik um: Es blickt auf die Entscheidungen, die nicht getroffen wurden, und auf die Rollen, die unbesetzt blieben.

Die vier Fälle dieses Kapitels sind öffentlich dokumentierte Vorfälle. Zu jedem existiert ein Gerichtsurteil, ein Behördenbescheid oder ein offizieller Untersuchungsbericht mit offen zugänglichen Materialien. Eines verbindet sie: Der technische Teil führte jeweils die Logik aus, die in ihn hineingebaut worden war — das Problem lag in der anthropologischen Schicht. Bei der Frage, wer hätte handeln können und müssen, als das System ein Ergebnis lieferte, das eine menschliche Reaktion verlangte; in der Rollenarchitektur, der Dokumentation, den Mandaten, den Verträgen. Die Technologie geriet nicht außer Kontrolle — ihr Kontext wurde nie fertig gebaut.

Jeder Fall: andere Branche, anderer Maßstab, andere Fehlerart. Aber dasselbe Muster: Die Rolle war unbesetzt, das Mandat unklar, die Dokumentation fehlend oder veraltet.

Norbert Wiener formuliert in *Cybernetics*, warum sich dieses Muster systematisch wiederholt: Ein automatisches System ohne geschlossene Rückkopplungsschleife zum Menschen erzeugt keine Fehler im engen Sinn, sondern *Drift* — eine langsame Abweichung von der Absicht, die niemand registriert, bis die kritische Störung eintritt. Stuart Russell übersetzt das in *Human Compatible* in die Gegenwart: Eskalation ist eine reguläre Phase des Regelkreises, kein Krisenereignis — und sie muss entworfen werden, statt darauf zu vertrauen, dass sie *sich schon ergibt*.

In jedem der vier Fälle lässt sich die konkrete Rolle benennen, die hätte existieren müssen und nicht existierte — und deren Fehlen die technisch korrekte Arbeit des Systems in einen Vorfall verwandelte. Fall A zeigt die fehlende *Eskalationsrolle*: ein Mensch mit operativem Mandat, die Automatik in einem konkreten Fehlermodus zu übersteuern. Fall B die fehlende *redaktionelle Rolle*: ein Modell-Kurator, der die Übernahme statistisch prädiktiver Merkmale aus dem kommerziellen Kontext in eine staatliche Entscheidung ohne gesonderte Begründung blockiert. Fall C den fehlenden *Compliance-Übersetzer*: ein Mensch, der die interne Modellentscheidung in eine Form bringt, die ein externer Prüfer ohne Offenlegung des Geschäftsgeheimnisses nachvollziehen kann. Fall D den fehlenden *Souveränitäts-Auditor*: eine Rolle, die die tatsächliche Unabhängigkeit vertraglicher Abhängigkeiten von geopolitischen Szenarien prüft. Vier Branchen, vier Fehlerarten — eine gemeinsame Struktur unbesetzter Rollen. Diese Struktur, nicht die Einzelgeschichten, ist der Gegenstand des Kapitels.

Fall A — Luftfahrt: Boeing 737 MAX MCAS (Lion Air JT610, 2018; Ethiopian ET302, 2019)

Kontext. Die Boeing 737 MAX ist eine Weiterentwicklung des Basismodells, in der zur Kompensation aerodynamischer Veränderungen durch die neuen Triebwerke das System MCAS (Maneuvering Characteristics Augmentation System) ergänzt wurde. In der Ausgangskonfiguration bezog MCAS seine Eingabe von einem einzigen Anstellwinkelsensor und drückte die Flugzeugnase automatisch nach unten, um einen Strömungsabriss zu verhindern.

Was geschah. Am 29. Oktober 2018 stürzte Flug Lion Air JT610 dreizehn Minuten nach dem Start in die Javasee: 189 Tote. Ein defekter Sensor lieferte einen falschen Anstellwinkel, MCAS drückte die Nase zyklisch nach unten. Am 10. März 2019 stürzte Flug Ethiopian Airlines ET302 im selben Szenario sechs Minuten nach dem Start ab: 157 Tote. Insgesamt 346 Opfer in fünf Monaten. Das weltweite Flugverbot für die gesamte 737-MAX-Flotte dauerte von März 2019 bis November 2020.

Was fehlte. Das operative Mandat zur manuellen Übersteuerung von MCAS — wann und wie der Pilot verpflichtet ist, die Automatik abzuschalten, welches Signal der Auslöser ist, wie die Kenntnismahme bestätigt wird — war in den Flughandbüchern nicht ausformuliert. MCAS wurde in der Dokumentation nur minimal erwähnt; die Schulung bestand aus einer Tablet-Einweisung von wenigen Dutzend Minuten, ohne das Szenario *aktives MCAS mit defektem Sensor* im Simulator. Die Abschalt-Schalter existierten physisch — aber ohne vorherige Vertrautheit mit diesem Fehlermodus hatten die Piloten keine trainierte Reaktion darauf.

Anthropologische Analyse. MCAS führte die einprogrammierte Logik aus — doch diese Logik, der Zertifizierungsprozess, die Konstruktionsdokumentation und die Pilotenausbildung erwiesen sich zusammen als unzureichend für den sicheren Betrieb. Nach dem Flugverbot genehmigte die FAA ein Korrekturpaket: Validierung über zwei Sensoren, Begrenzung der Reaktivierung, verpflichtendes Simulatortraining für alle 737-MAX-Piloten und ein explizites Protokoll zur manuellen Übersteuerung. Der Untersuchungsausschuss des US-Repräsentantenhauses legte die Verantwortung 2020 auf die Kombination aus Boeings Konstruktionsentscheidungen, Zertifizierungsmängeln und der Aufsicht der FAA. Die Eskalationsrolle — der Pilot als Träger des Mandats zur manuellen Übersteuerung in einem konkreten Fehlermodus — war weder durch Training noch durch ein explizites Protokoll abgesichert; und das war nur eine der Vakanzen, die sich bis zur Katastrophe aufschichteten.[¹]

Fall B — öffentlicher Sektor: der AMS-Algorithmus (Arbeitsmarktservice Österreich, 2018–2020)

Kontext. Das Arbeitsmarktservice (AMS) ist die staatliche Arbeitsvermittlung Österreichs. 2018 führte das AMS einen Algorithmus ein, der die Chancen Arbeitsloser auf Rückkehr in den Arbeitsmarkt statistisch bewerten und sie in drei Gruppen einteilen sollte: A (hohe Chancen), B (mittlere), C (niedrige). Entwickelt wurde er von der Beratungsfirma Synthesis Forschung, Projektkosten rund 240 000 Euro. Die Gruppenzuteilung sollte in Entscheidungen über den Zugang zu Qualifizierungsprogrammen einfließen.

Was geschah. Das erste öffentliche Audit des Modells (Algorithm-Watch, 2019) zeigte: Die Formel vergab negative Gewichte für die Merkmale *weiblich*, *betreuungspflichtige Kinder* (nur bei Frauen, nicht bei Männern), *Alter 30+*, *Alter 50+*, *gesundheitliche Beeinträchtigung*, *Nicht-EU-Staatsbürgerschaft*. Eine Frau mit Kind erhielt automatisch einen niedrigeren Wert als ein Mann mit identischem Lebenslauf. Das AMS begründete das als *Abbild der realen Chancen am Markt*. Am 20. August 2020 erklärte die Datenschutzbehörde (DSB) das System per Bescheid D213.1020/2020-0513.605 für unvereinbar mit dem Datenschutzrecht: fehlende Rechtsgrundlage, fehlende Garantien gegen Diskriminierung, de facto automatisierte Entscheidung trotz deklarierter *bloßer Unterstützung*. Der Einsatz wurde für die Dauer der Anfechtung ausgesetzt. 2025 hob das Bundesverwaltungsgericht (BVwG) den DSB-Bescheid auf — das Gericht kam zum Schluss, dass die Datenverarbeitung des AMS in diesem Format nicht gegen die DSGVO verstößt und die Bewertung des Diskriminierungseffekts nicht in die Zuständigkeit der DSB fällt. Gesellschaftlich und wissenschaftlich bleibt die Frage offen; juristisch wurde das System nie endgültig für rechtswidrig erklärt.

Was fehlte. Vor dem Start des Modells stellte niemand die Frage: *Welche Korrelationen aus historischen Arbeitsmarktdaten wollen wir nicht in eine staatliche Entscheidung über Programmzugänge übernehmen?* Das Team behandelte die Aufgabe als technische (künftige Chancen aus vergangenen vorhersagen), nicht als administrative (eine begrenzte Ressource gerecht verteilen). Die Rolle des Modell-Kurators im staatlichen Kontext — mit dem Mandat, Merkmale zu blockieren, die mit geschützten Kategorien korrelieren, auch wenn sie statistisch prädiktiv sind — war institutionell nicht verankert.

Anthropologische Analyse. Das Modell war für die Prognose der Rückkehrwahrscheinlichkeit auf Basis historischer Daten gebaut und leistete genau das, wofür es entwickelt wurde. Die redaktionelle Rolle über den Merkmalen — mit dem Mandat zu sagen: *dieses Merkmal ist als Prognosemerkmal im kommerziellen Kontext zulässig, braucht aber als Grundlage einer staatlichen Entscheidung eine eigene Begründung* — blieb unbesetzt. Das Projekt kostete die Steuerzahler rund 240 000 Euro, produzierte einen öffentlichen Skandal, den DSB-Bescheid von 2020 und dessen Aufhebung durch das BVwG 2025. Unabhängig vom juristischen Endergebnis bleibt der Fall ein Lehrbuchbeispiel dafür, wie eine administrative Aufgabe (Verteilung einer begrenzten Ressource) als technische (Chancen vorhersagen) verpackt werden kann — unter Auslassung der anthropologischen Schicht.[²]

Fall C — Kredit-Scoring: Schufa, EuGH C-634/21 (Urteil vom 7. Dezember 2023)

Kontext. Die Schufa Holding AG (Wiesbaden) ist die größte deutsche Kreditauskunftei. Ihr automatisiertes Scoring — der Schufa-Score — wird von Banken, Leasing- und Telekommunikationsunternehmen für Entscheidungen über Kredite, Mietverhältnisse und Mobilfunkverträge herangezogen. Der Algorithmus zur Bildung des Scores wurde nie veröffentlicht: Die Schufa beruft sich auf das Geschäftsgeheimnis.

Was geschah. Einer Bürgerin in Hessen wurde ein Kredit unter Verweis auf einen niedrigen Schufa-Score verweigert. Sie verlangte eine Erklärung der Gewichte und Faktoren. Die Schufa teilte nur den Endwert und eine allgemeine Methodenbeschreibung mit und verweigerte die Offenlegung der internen Parameter. Das Verwaltungs-

gericht Wiesbaden legte die Frage dem Gerichtshof der Europäischen Union vor. Mit dem Urteil in der Rechtssache C-634/21 (7. Dezember 2023) stellte der EuGH fest: Stützt sich ein Dritter maßgeblich auf ein automatisiertes Scoring, wenn er über einen Menschen entscheidet, fällt dieses Scoring unter Artikel 22 DSGVO – mit zwingenden Garantien (menschliche Überprüfung, aussagekräftige Erklärung, Widerspruchsrecht). Das formale *wir berücksichtigen den Score nur* der Bank nimmt die Entscheidung nicht aus Artikel 22 heraus, wenn auf einen niedrigen Score faktisch fast immer die Ablehnung folgt. Nach dem Urteil überarbeiteten Kreditauskunfteien in der EU ihre Transparenz- und Widerspruchspraxis.

Was fehlte. Das System klassifizierte nach den internen Spezifikationen, auf denen sein Geschäftsmodell aufbaut. Die Schnittstelle zwischen ihm und dem externen Prüfer – erst der Kundin, dann dem nationalen Gericht, dann dem EuGH – war nie entworfen worden. Kein Werkzeug, das die Modellentscheidung einem nicht-technischen Prüfer erklärt. Kein Mechanismus, der das Scoring-Ergebnis aussagekräftig begründet, ohne das Geschäftsgeheimnis offenzulegen.

Anthropologische Analyse. Das Scoring arbeitete innerhalb seiner Geschäftslogik. Die Rolle des Compliance-Übersetzers – eines Menschen, der die interne Modellentscheidung in eine Form bringt, die ein externer Prüfer ohne Zugriff auf das Modell verifizieren kann – existierte weder bei der Schufa noch bei den meisten vergleichbaren Auskunfteien. Ohne diese Übersetzung landet automatisiertes Scoring, sobald Dritte sich maßgeblich darauf stützen, in der Falle des Artikels 22 – nicht wegen der Automatisierung selbst, sondern wegen der fehlenden Brücke zwischen Modell und Mensch mit Kontrollmandat. Die klassisch fehlende KI-Operator-Rolle in der Compliance-Dimension.[³]

Fall D — Cloud-Sanktionen 2022: Dienst einschränkungen bei AWS, Microsoft, Google

Kontext. Europäische Industrie- und Finanzunternehmen betreiben Teile ihrer Machine-Learning-Pipelines bei öffentlichen Cloud-Anbietern (AWS, Microsoft Azure, Google Cloud) — wegen Skaleneffekten und des Zugangs zu verwalteten KI- und ML-Diensten. Viele von ihnen hatten Tochtergesellschaften in Russland und Belarus oder Partner in sanktionierten Jurisdiktionen.

Was geschah. Am 4. März 2022 kündigte Microsoft an, den Neukauf von Produkten und Diensten in Russland auszusetzen. Am 8. März 2022 teilte AWS mit, keine neuen Kunden in Russland und Belarus mehr zu registrieren. Google stoppte neue Abonnements kostenpflichtiger Google-Cloud-Dienste. Im Dezember 2023 erweiterte das zwölfte EU-Sanktionspaket (Beschluss (GASP) 2023/2874 und Verordnung (EU) 2023/2878 des Rates vom 18. Dezember 2023) die Beschränkungen auf die Lieferung von Unternehmenssoftware an in Russland registrierte juristische Personen. In der Folge musste eine Reihe europäischer Unternehmen mit gemeinsamer Infrastruktur ihre internationalen Informationssysteme und die Zugriffsverfahren auf Cloud-Infrastruktur umbauen. Öffentliche technische Aufarbeitungen solcher Umbauten existieren kaum — deshalb zählt dieser Fall nicht als Geschichte eines einzelnen Unternehmens, sondern als Beispiel einer neuen Risikoklasse, die mit den Sanktionsentscheidungen von 2022 sichtbar wurde.

Was fehlte. Bis 2022 umfasste die Standard-Vertragsprüfung für Cloud-Dienste die Verfügbarkeitszusagen (SLA — Service Level Agreement: garantierte Verfügbarkeits- und Reaktionszeiten), den Preis, die technische Kompatibilität und die DSGVO-Konformität. Das Szenario einer geopolitisch motivierten Zugangssperre — nicht

durch technisches Versagen, sondern durch die rechtliche Verpflichtung des Anbieters gegenüber seinem Staat — war bei den meisten europäischen Unternehmen weder durch ein Drehbuch noch durch einen Notfallplan noch durch eine Ausstiegsstrategie abgedeckt.

Anthropologische Analyse. Technisch arbeiteten die Systeme bis zum Moment der vertraglichen oder rechtlichen Einschränkung korrekt. Die Rolle des Souveränitäts-Auditors — eines Menschen, der die tatsächliche Unabhängigkeit vertraglicher Abhängigkeiten von geopolitischen Szenarien prüft und einen Notfallplan für die plötzliche Abschaltung baut — war nicht besetzt. Der Einkauf wählte nach Preis und SLA; die IT-Architektur nach technischer Kompatibilität; die Juristen nach Standardklauseln. Keine dieser Abteilungen hatte das Mandat für die Frage: *Unter welchen Szenarien kann dieser Anbieter das Recht verlieren, uns zu bedienen — und was dann?*^[4]

Das Muster: In allen vier Fällen war die Anthropologie das Problem, nicht das Modell

Systematisiert man die vier Fälle:

Fall A (Boeing MCAS). Das Mandat zur manuellen Übersteuerung der Automatik im Fehlermodus war nicht operativ. Technisch führte MCAS die einprogrammierte Logik aus — doch diese Logik, die Zertifizierung und die Dokumentation entstanden ohne explizites Protokoll für das Pilotenhandeln im konkreten Fehlermodus. Die Eskalationsrolle war weder durch Training noch durch ein Protokoll abgesichert.

Fall B (AMS-Algorithmus). Ein Merkmals-Kurator mit dem Mandat zu sagen *diese Variable ist im kommerziellen Kontext zulässig, braucht aber in einer staatlichen Entscheidung eine eigene Begründung* — fehlte. Technisch leistete das Modell die Prognose, für die es gebaut war. Die redaktio-

nelle Rolle über den Merkmalen im administrativen Kontext blieb vakant. Die juristische Bewertung wurde angefochten und drehte sich; die anthropologische Lektion blieb.

Fall C (Schufa, EuGH). Der Mechanismus der Compliance-Übersetzung für externe Prüfer fehlte. Technisch arbeitete das Scoring nach seiner internen Geschäftslogik. Es gab keine Brücke zwischen Modell und Mensch mit Kontrollmandat (Kundin, Regulierer, Gericht) — und genau das macht automatisiertes Scoring unter bestimmten Bedingungen zum Fall für Artikel 22 DSGVO.

Fall D (Cloud-Sanktionen 2022). Das Souveränitäts-Audit der vertraglichen Abhängigkeiten fehlte. Technisch erfüllten die Systeme ihre Funktion bis zur rechtlichen Einschränkung durch den Anbieter. Die Haftungsrolle hatte das geopolitische Risiko in den Anbieterabhängigkeiten nicht identifiziert.

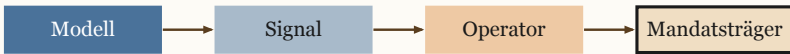
In allen vier Fällen tat das Modell, wofür es gebaut war — oder versuchte es. Das Problem entstand dort, wo die menschliche Struktur — Rollen, Mandate, Dokumentation, Verträge — keinen Kontext für die Reaktion auf das Systemergebnis bereitstellte.

Der Eskalationsfluss — und wo er abreißt

Obere Ebene: wie die Organisation sich Eskalation vorstellt.

Untere Ebene: wie sie tatsächlich läuft.

Erwartungsebene — „wir haben Eskalation“



Realitätsebene — wo der Eskalationsfluss abreißt



Fall B — Ebene Signal: Daten-Kurator fehlt
die Trainingsrolle hat kein redaktionelles Mandat

Fall C — Ebene Operator: Übersetzung für den Prüfer fehlt
der Operator kann extern nichts belegen

Fall A — Ebene Mandatsträger: Mandat zum Eingriff fehlt
kein Training für den Fehlermodus

Vertragliche Abhängigkeiten vom Anbieter

Fall D — Souveränitäts-Auditor fehlt
die Sanktions-Abschaltung war nicht vorgesehen
die Haftungsrolle hat das Anbieter-Risiko nicht identifiziert

Kein Abriss geschah im Modell. Jeder an einer unbesetzten Rolle.

Ankerdiagramm K6/1. Vier Eskalationsfälle A–D: Kein Abriss geschah im Modell — jeder an einer unbesetzten menschlichen Rolle auf einer konkreten Ebene.

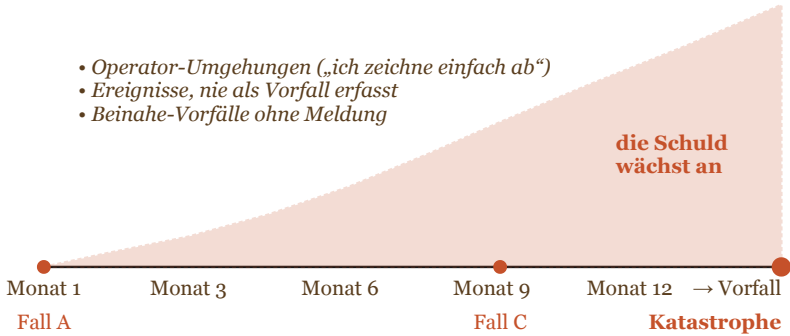
Genau das ist der Sinn des anthropologischen Ansatzes: nicht den technischen Blick zu ersetzen, sondern ihn um Fragen zu ergänzen, die die technische Analyse nicht stellt.

Die unsichtbare Eskalationsschuld

unsichtbare Schuld, die sich zwischen den Vorfällen ansammelt

„Eskalationsschuld“ — Umgehungen, die niemand notiert hat

- Operator-Umgehungen („ich zeichne einfach ab“)
- Ereignisse, nie als Vorfall erfasst
- Beinahe-Vorfälle ohne Meldung



Sichtbare Vorfälle sind Punkte.

Die unsichtbare Schuld ist die Fläche unter der Linie.

ERWARTETE VS. REALE KONTROLLE

wiederkehrendes Motiv (K6 · K8 · K16 · K21)

Erwartungsebene

Mensch im Regelkreis · der Experte prüft · die Führungskraft zeichnet ab



Realitätsebene

Bescheinigung statt Prüfpfad · Unterschrift ohne Prüfung · Mandat ohne Hebel

*Kontrolle existiert, wo ein Mensch Mandat, Kanal und Zeit hat —
nicht, wo ein Name eingetragen ist.*

Drei Fragen für den eigenen Kontext

1. **Mandate.** Existieren für jede Eskalationsstufe Ihres Systems schriftlich benannte Personen mit klarem Entscheidungsmandat — auch außerhalb der Bürozeiten?
2. **Daten-Kuration.** Wer prüft Ihre Trainings- und Nachtrainings-Datensätze auf Anomalien, die das Modell *nicht* lernen soll?
3. **Anbieter-Souveränität.** Welche Ihrer KI-Komponenten hängen von einem Anbieter ab, der in einem geopolitischen Szenario rechtlich gezwungen sein könnte, den Zugang einzuschränken — und ist dieses Szenario in einem Notfallplan abgedeckt?

Quellen

1. FAA, „Summary of the FAA's Review of the Boeing 737 MAX“, 18. November 2020; US House Committee on Transportation and Infrastructure, „Final Committee Report on the Design, Development, and Certification of the Boeing 737 MAX“, 15. September 2020.
2. Nicolas Kayser-Bril, „Austria's employment agency rolls out discriminatory algorithm, sees no problem“, AlgorithmWatch, 6. Oktober 2019; Datenschutzbehörde Österreich, Bescheid D213.1020/2020-0513.605 vom 20. August 2020; Bundesverwaltungsgericht Österreich, Aufhebung des DSB-Bescheids, 2025.
3. Urteil des Gerichtshofs (Erste Kammer) vom 7. Dezember 2023, Rechtssache C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957.
4. Brad Smith, „Microsoft suspends new sales in Russia“, Microsoft On the Issues Blog, 4. März 2022; Amazon Staff, „Amazon's assistance in Ukraine“, aboutamazon.com, 8. März 2022; Beschluss (GASP)

2023/2874 des Rates vom 18. Dezember 2023; Verordnung (EU)
2023/2878 des Rates vom 18. Dezember 2023 zur Änderung der Ver-
ordnung (EU) Nr. 833/2014, EUR-Lex.