Public AI Failure Patterns and WEKID™ Governance Responses

Executive Evidence Edition

Lessons from 45 Real-World Incidents Across Government and Industry

June 2026

Version 1.0

Prepared by WEKID™ LLC

© 2025–2026 WEKID LLC. All rights reserved.

Trademarks. WEKID™ and the WEKID™ epistemic framework are trademarks of WEKID LLC. The absence of a trademark symbol in any instance does not constitute a waiver of these rights. All other trademarks, service marks, product names, company names, and organizational references remain the property of their respective owners and are used solely for identification, commentary, educational analysis, and governance discussion.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Copyright and Trademark Notice

© 2025–2026 WEKID LLC. All rights reserved.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WEKID™ and the WEKID™ epistemic framework are trademarks of WEKID LLC. The absence of a trademark symbol in any instance does not constitute a waiver of these rights.

All other trademarks, service marks, product names, company names, and organizational references remain the property of their respective owners and are used solely for identification, commentary, educational analysis, and governance discussion.

Interpretive Use Notice

The incidents described throughout this publication are drawn from publicly reported events, investigations, court filings, regulatory actions, academic studies, government reports, and reputable media sources available at the time of publication.

WEKID™ was not deployed, licensed, implemented, evaluated, or otherwise adopted by any of the organizations, agencies, institutions, or individuals referenced in these incidents prior to the events described.

Accordingly, nothing in this publication should be interpreted as suggesting that WEKID™ was involved in, contributed to, endorsed, approved, or failed to prevent any of the outcomes discussed.

The analyses presented herein are retrospective and illustrative. They represent an application of the WEKID™ epistemic framework to publicly documented incidents in order to identify recurring patterns of governance failure and demonstrate how alternative controls, escalation mechanisms, and evaluation practices might have altered the decision pathway.

These analyses do not constitute factual determinations regarding causation, negligence, liability, intent, legal responsibility, regulatory compliance, or wrongdoing.

No framework can eliminate the possibility of failure.

The purpose of this publication is not to claim that WEKID™ would have prevented every incident described.

Rather, its purpose is to demonstrate how recurring failure patterns can be made more visible, more traceable, and more governable before harm occurs.

The question this brief invites leaders to consider is not whether past failures can be changed.

It is whether future failures can be recognized and governed before they become public crises.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Executive Summary

Artificial intelligence failures have become increasingly visible across nearly every sector of society.

Healthcare systems have produced unsafe recommendations. Financial institutions have struggled to explain algorithmic decisions affecting consumers. Educational institutions have faced backlash when automating moments of human significance. Government agencies have incorrectly denied benefits and applied probabilistic models to decisions affecting civil liberties. Military organizations now confront the possibility of autonomous systems exercising increasingly consequential forms of authority.

Viewed individually, these incidents appear unrelated. Viewed collectively, a different pattern emerges.

Across industries, organizations repeatedly grant authority to intelligent systems beyond the quality of evidence, context, experience, and judgment available to support their decisions.

These are not simply failures of intelligence. They are failures of governance.

The purpose of this Executive Evidence Edition is not to criticize organizations or second-guess decisions made under complex circumstances. Rather, it is to reveal recurring conditions under which intelligent systems fail when granted authority beyond their demonstrated competence.

The forty-five cases presented throughout this publication represent publicly documented examples drawn from healthcare, education, finance, legal practice, government services, defense, transportation, technology, and commercial enterprise.

Each case examines four questions:

1. What underlying failure pattern emerged?
2. How did that pattern manifest in practice?
3. Why should leaders care?
4. How would WEKID™ interpret the governance response?

The industries differ. The technologies differ. The names differ. The governance failures repeat.

Organizations that govern intelligence deliberately will recognize these signals early, intervene appropriately, and preserve trust before failure becomes public.

Organizations that do not may discover that AI failures are rarely technical surprises. They are governance surprises.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

How to Read This Brief

The cases in this publication are intentionally organized around recurring patterns of failure rather than by industry alone. This reflects a core observation underlying the WEKID™ framework: the structure of intelligence remains constant even as technologies evolve.

Healthcare, education, defense, finance, and government administration appear distinct. Yet the same epistemic failures emerge repeatedly across all of them.

Each case follows a common format:

- Failure Pattern — the underlying epistemic mistake.
- The Incident — what happened and how the failure unfolded.
- The Impact — the consequences that followed.
- WEKID™ Governance Interpretation — how the WEKID framework interprets the breakdown.
- Executive Reflection — questions leaders should ask within their own institutions.

The purpose of these cases is not retrospective judgment. It is prospective governance. The value of failure patterns lies not in explaining the past. It lies in recognizing the future before it arrives.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Contents

Copyright and Trademark Notice 2
Interpretive Use Notice 2
Executive Summary 3
How to Read This Brief 4
Contents 5
Index of Cases 7
The Cases 9
WF-001 — Fabricated Legal Citations in Court Filings 10
WF-002 — AI-Generated Medical Advice Without Risk Qualification 12
WF-003 — Automated Hiring and Screening Bias 14
WF-004 — Autonomous Trading and Market Disruption 16
WF-005 — Government Benefits Eligibility Errors 18
WF-006 — Predictive Policing and Risk Scoring 20
WF-007 — Content Moderation Errors at Scale 22
WF-008 — Hallucinated Financial Analysis 24
WF-009 — Autonomous Infrastructure Failures (AIOps) 26
WF-010 — Inability to Explain AI Decisions After Incidents 28
WF-011 — Executive Surprise and Board-Level Blindness 30
WF-012 — Workforce Over-Trust and Skill Atrophy 32
WF-013 — Chatbot Hallucinated Policy Presented as Authoritative Guidance 34
WF-014 — Fluent Expertise Mistaken for Clinical Competence 36
WF-015 — Correct Predictions Applied Beyond Their Authority 38
WF-016 — Unbounded Learning from Hostile Environments 39
WF-017 — Corrective Constraints Creating New Distortions 40
WF-018 — Automation Applied Without Procedural Fairness 42
WF-019 — Automated Suspicion Without Appeal 44
WF-020 — Optimization Without Human Context 46
WF-021 — Authenticity Collapse (Synthetic Media) 48
WF-022 — Automation Overriding Operator Understanding 50
WF-023 — AI Name-Announcement Failure at Commencement 52
WF-024 — Automated Ceremony Recognition Rejected as Inappropriate 54
WF-025 — Stakeholder Rejection of AI in Ceremonial Roles 56
WF-026 — Institutional Empathy Outsourced to AI 58
WF-027 — AI Output Treated as Evidence of Cheating 60

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-028 — AI-Detection Scores Mistaken for Verdicts 62
WF-029 — Scaling Student Support Beyond Demonstrated Competence 64
WF-030 — AI Targeting Scores Treated as Lethal Judgment 66
WF-031 — Battlefield Autonomy Expanding Faster Than Governance 68
WF-032 — Escalatory Autonomy Arms Race 70
WF-033 — Disputed Autonomous Lethal Engagement 72
WF-034 — Decision Support Drifting Toward Decision Authority 74
WF-035 — Scaling Autonomy Before Explainability Matures 76
WF-038 — Overreliance on a Flawed Prediction Model 77
WF-039 — Predictive Confidence Mistaken for Treatment Authority 79
WF-040 — AI-Assisted Procedure Causing Surgical Harm 81
WF-042 — Diagnosis Without Adequate Evidence Gathering 83
WF-043 — Clinicians as the Liability Sink for AI Error 85
WF-045 — Automated Workflow Disruption in High-Stakes Care 86
WF-046 — Probabilities Replacing Clinical Judgment in Coverage 88
WF-047 — False Positives in Security Screening 89
WF-050 — Algorithmic Denial of Post-Acute Care 91
WF-053 — Overstated Autonomy and Degraded Human Vigilance 93

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Index of Cases

All forty-five cases appear below in catalogue order, with sector and the governance outcome WEKID™ would produce. The full entries follow in the same order.

ID	Case	Sector	Decision outcome
WF-001	Fabricated Legal Citations in Court Filings	Legal	Rejected
WF-002	AI-Generated Medical Advice Without Risk Qualification	Healthcare	Constrained
WF-003	Automated Hiring and Screening Bias	Employment / HR	Constrained / Remediation
WF-004	Autonomous Trading and Market Disruption	Financial markets	Constrained
WF-005	Government Benefits Eligibility Errors	Public sector	Constrained
WF-006	Predictive Policing and Risk Scoring	Criminal justice	Constrained / Rejected
WF-007	Content Moderation Errors at Scale	Online platforms	Constrained
WF-008	Hallucinated Financial Analysis	Finance / media	Remediation / Rejected
WF-009	Autonomous Infrastructure Failures (AIOps)	Cloud infrastructure	Constrained
WF-010	Inability to Explain AI Decisions After Incidents	Financial services	Remediation Required
WF-011	Executive Surprise and Board-Level Blindness	Professional services	Approved w/ Monitoring
WF-012	Workforce Over-Trust and Skill Atrophy	Cross-industry	Remediation Required
WF-013	Chatbot Hallucinated Policy Presented as Authoritative Guidance	Aviation	Rejected
WF-014	Fluent Expertise Mistaken for Clinical Competence	Healthcare	Constrained
WF-015	Correct Predictions Applied Beyond Their Authority	Real estate	Constrained
WF-016	Unbounded Learning from Hostile Environments	Consumer AI	Rejected
WF-017	Corrective Constraints Creating New Distortions	Generative AI	Remediation Required
WF-018	Automation Applied Without Procedural Fairness	Public sector	Constrained / Rejected
WF-019	Automated Suspicion Without Appeal	Public sector	Rejected
WF-020	Optimization Without Human Context	Logistics	Constrained
WF-021	Authenticity Collapse (Synthetic Media)	Elections	Rejected
WF-022	Automation Overriding Operator Understanding	Aviation	Rejected
WF-023	AI Name-Announcement Failure at Commencement	Education	Constrained
WF-024	Automated Ceremony Recognition Rejected as Inappropriate	Education	Constrained
WF-025	Stakeholder Rejection of AI in Ceremonial Roles	Education	Constrained
WF-026	Institutional Empathy Outsourced to AI	Education	Constrained
WF-027	AI Output Treated as Evidence of Cheating	Education	Rejected
WF-028	AI-Detection Scores Mistaken for Verdicts	Education	Constrained

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

ID	Case	Sector	Decision outcome
WF-029	Scaling Student Support Beyond Demonstrated Competence	Education	Constrained
WF-030	AI Targeting Scores Treated as Lethal Judgment	Defense & military	Rejected
WF-031	Battlefield Autonomy Expanding Faster Than Governance	Defense & military	Constrained
WF-032	Escalatory Autonomy Arms Race	Defense & military	Constrained
WF-033	Disputed Autonomous Lethal Engagement	Defense & military	Rejected
WF-034	Decision Support Drifting Toward Decision Authority	Defense & military	Constrained
WF-035	Scaling Autonomy Before Explainability Matures	Defense & military	Constrained
WF-038	Overreliance on a Flawed Prediction Model	Healthcare	Remediation Required
WF-039	Predictive Confidence Mistaken for Treatment Authority	Healthcare	Constrained
WF-040	AI-Assisted Procedure Causing Surgical Harm	Healthcare	Remediation Required
WF-042	Diagnosis Without Adequate Evidence Gathering	Healthcare	Constrained
WF-043	Clinicians as the Liability Sink for AI Error	Healthcare	Remediation Required
WF-045	Automated Workflow Disruption in High-Stakes Care	Government & healthcare	Constrained
WF-046	Probabilities Replacing Clinical Judgment in Coverage	Insurance & healthcare	Constrained
WF-047	False Positives in Security Screening	Homeland security	Remediation Required
WF-050	Algorithmic Denial of Post-Acute Care	Insurance & healthcare	Constrained
WF-053	Overstated Autonomy and Degraded Human Vigilance	Transportation	Constrained

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

The Cases

The brief opens with a fully worked example (WF-001) showing the analytical format in depth. The remaining cases follow the same structure, drawn from the case database.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-001 — Fabricated Legal Citations in Court Filings

Failure pattern	Hallucinated facts treated as authoritative truth.
Primary WEKID™ layer(s)	Primary: Data · Secondary: Wisdom
Domain	Legal practice · Professional services · Judicial proceedings
Period	2023
Governance trigger	Fabrication Gate (fabricated citations, sources, or tool outputs detected); Data Integrity Gate (unverified claims in high-stakes outputs)
Decision-Matrix outcome	Rejected — outputs containing fabricated legal authorities should never advance to filing; mandatory escalation and verification required
Organizations	Mata v. Avianca · OpenAI (ChatGPT)

The Incident. In early 2023, attorneys representing plaintiff Roberto Mata filed legal briefs in the United States District Court for the Southern District of New York in a case against Avianca Airlines involving alleged injuries sustained from a serving cart during an international flight. Seeking precedent to oppose Avianca's motion to dismiss, attorney Steven Schwartz turned to ChatGPT for legal research assistance. The system generated multiple cases that appeared authentic — plausible case names, judicial reasoning, docket references, reporter citations, quotations, and summaries of holdings. Among them were Varghese v. China Southern Airlines, Martinez v. Delta Air Lines, Miller v. United Airlines, Petersen v. Iran Air, Shaboon v. EgyptAir, and Zicherman v. Korean Air Lines.

The cases did not exist. When opposing counsel could not locate the authorities in Westlaw, LexisNexis, PACER, or traditional reporters, Avianca's attorneys alerted the court. Judge P. Kevin Castel ordered the plaintiff's attorneys to provide copies of the cited opinions. Schwartz returned to ChatGPT, which reassured him the cases were genuine and even generated excerpts from the fictitious opinions, which were then submitted to the court. Only after the court intensified its inquiry did the attorneys realize the cases had never existed. In June 2023, Judge Castel issued sanctions, observing that while there is nothing improper about using a reliable AI tool, existing rules place a gatekeeping duty on attorneys to ensure the accuracy of their filings.

The Impact. Professional consequences included public embarrassment, reputational damage, judicial sanctions, and mandatory client/court notifications. Institutionally, courts across the U.S. began issuing AI-use guidance, several jurisdictions requiring attorneys to certify human review and independent verification of authorities. Culturally, the incident challenged assumptions about expertise: if lawyers trained to verify precedent could be persuaded by fabricated citations, the same risk applies to professions with weaker validation cultures.

WEKID™ Governance Interpretation. Although superficially a Data failure (the citations were false), the deeper breakdown was at the Wisdom layer. Multiple intervention opportunities existed — opposing counsel questioned the authorities, the court requested copies, verification systems failed to confirm — yet confidence persisted because fluency created legitimacy. The presentation of expertise substituted for expertise itself.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WEKID layer	Question	Answer
Data	Are the facts real?	No. Fabrication Gate activated.
Information	Is the contextual evidence complete?	No independent legal validation occurred.
Knowledge	Are the inferences supported?	No. Reasoning relied on nonexistent precedent.
Wisdom	Should this output ever have been trusted without escalation?	Absolutely not — high-stakes content capable of affecting judicial outcomes.

Executive Reflection

Leaders should ask:

1. **Where do fabricated outputs have pathways into authoritative decisions?**
   Could similar failures enter contracts, board materials, regulatory submissions, public communications, or compliance reports?
2. **What verification mechanisms exist before AI-generated outputs become institutional truth?**
   Are they mandatory, or optional?
3. **Have confidence and fluency become proxies for correctness?**
   How would your organization know the difference?

Documented Examples. Mata v. Avianca, Inc. — S.D.N.Y. Individuals: Roberto Mata; Steven A. Schwartz; Peter LoDuca; Judge P. Kevin Castel. Key dates: March 2023 brief submitted; May 2023 fabrications discovered; June 8, 2023 sanctions issued.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-002 — AI-Generated Medical Advice Without Risk Qualification

Failure pattern	Correct information delivered without judgment.
Domain	Healthcare · Mental health
Period	2022–2023
Governance trigger	High-Stakes Wisdom Gate (cap 60/100) — uncertainty disclosure and escalation required
Decision-Matrix outcome	Constrained
Organizations	NEDA · Cass (chatbot vendor) — "Tessa"

The Incident. The National Eating Disorders Association had run a volunteer-staffed helpline for roughly 24 years, handling about 70,000 contacts in 2022. In spring 2023 NEDA moved to close the human helpline and direct users to "Tessa," a chatbot running a guided program called Body Positive. Tessa had quietly launched in February 2022 — before the post-ChatGPT wave — and NEDA repeatedly stressed it was a pre-scripted bot, not a generative GPT system. It was built and operated by the vendor Cass (CEO Michiel Rauws), with clinical content designed by researcher Dr. Ellen Fitzsimmons-Craft. Cass attributed the later behavior to a 2022 "systems upgrade" that added an enhanced question-and-answer capability beyond the original scripted responses.

Users reported that Tessa recommended calorie counting, weight loss, and body-fat/skin-fold measurement — advice contraindicated for people with eating disorders — even after a user disclosed an eating disorder. An early warning had come in October 2022, when Monika Ostroff (executive director of the Multi-Service Eating Disorders Association) sent NEDA screenshots of the bot steering her toward "healthy" snacks and away from "unhealthy" foods. The issue became public in late May 2023 when activist Sharon Maxwell described her interactions; she reported that NEDA initially posted that her claims were false, then deleted that post after she supplied screenshots. Psychologist Dr. Alexis Conason independently reproduced the harmful outputs. NEDA took Tessa offline around May 30–June 1, 2023, acknowledging it "may have given information that was harmful"; CEO Liz Thompson at one point attributed part of the failure to "bad actors." The outputs were not wrong in a narrow factual sense — they were unsafe in practice, lacking uncertainty and escalation to professional care.

The Impact. Public backlash, suspension of the chatbot, regulatory scrutiny of digital health tools, and broader concern about deploying consumer AI into clinical contexts. The episode drew added scrutiny because the helpline closure followed staff moves to unionize, and Tessa saw a reported ~600% traffic surge after the closure announcement.

WEKID™ Governance Interpretation. A Wisdom gate enforces uncertainty disclosure and escalation paths in high-stakes contexts via uncertainty-disclosure requirements, risk thresholds, and policy-alignment checks (uncertainty-calibration score; risk-flag rate). Data-layer correctness is necessary but insufficient: when the cost of error is serious, the system must communicate the limits of its reliability and route users toward human judgment. The High-Stakes Wisdom Gate triggers when risk language is absent.

Executive Reflection

Leaders should ask:

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

4. **Where do our AI systems deliver factually plausible answers in contexts where the cost of being wrong is human harm?**
   Do those systems disclose uncertainty and route users to qualified humans?
5. **When we replace human judgment with automation, do we preserve a path to escalation?**
   Who is accountable when the safe answer is "I cannot help with this"?
6. **Do we treat domain correctness as sufficient evidence of safety?**
   How would we detect advice that is accurate in isolation but harmful in context?

Documented Examples. NEDA "Tessa" chatbot, running the Body Positive program. Launched February 2022; early warning October 2022 (Monika Ostroff); harmful advice publicized late May 2023 (Sharon Maxwell; reproduced by Dr. Alexis Conason); suspended ~June 1, 2023. NEDA CEO Liz Thompson; chatbot vendor Cass (CEO Michiel Rauws); content designer Dr. Ellen Fitzsimmons-Craft.

Sources. CNN Business (Jun 1, 2023); NPR Shots (Jun 8, 2023); Fortune (May 31, 2023); BBC News; NBC News.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-003 — Automated Hiring and Screening Bias

Failure pattern	Misinterpreted historical data embedded in decision logic.
Domain	Employment / HR
Period	2014–2023
Governance trigger	Domain bias gate; Knowledge deficit caps autonomy
Decision-Matrix outcome	Constrained / Remediation
Organizations	Amazon · iTutorGroup

The Incident. Resume-screening and candidate-scoring systems systematically disadvantaged applicants from certain groups. The bias was not deliberate; it was learned from historical hiring data reflecting past inclusion and exclusion. The systems appeared neutral; the outcomes were not.

Amazon

Amazon began building an experimental recruiting tool in 2014 to score candidates on a 1-to-5 star scale, training it on roughly ten years of past résumés drawn from a predominantly male applicant pool. By 2015 the team found it was not rating technical candidates in a gender-neutral way: it penalized résumés containing the word "women's" (as in "women's chess club captain"), downgraded graduates of two all-women's colleges, and favored aggressive verbs such as "executed" and "captured" that appeared more often on male résumés. Amazon edited those specific terms but lost confidence the model was neutral elsewhere and scrapped the project around 2017. Reuters (Jeffrey Dastin) reported the story in October 2018. Amazon said the tool was never the sole basis for hiring and was used only in a trial phase.

iTutorGroup

iTutorGroup (comprising Shanghai Ping'An Intelligent Education Technology Co. and Tutor Group Limited) hired U.S.-based remote tutors to teach English to students in China. Its application software was programmed to automatically reject female applicants aged 55 or older and male applicants aged 60 or older, screening out more than 200 qualified applicants in 2020 — a violation of the Age Discrimination in Employment Act. The behavior surfaced when an applicant submitted two otherwise-identical applications differing only in birth date: the application with the older birth date was rejected, while the one with a more recent birth date received an interview.

The Impact. Regulatory investigations, withdrawal of automated tools, reputational damage, and renewed legal scrutiny of algorithmic employment decisions. The EEOC's case against iTutorGroup became the agency's first AI-hiring-discrimination settlement.

WEKID™ Governance Interpretation. Knowledge-layer evaluation detects improper generalization — inferring future fitness from past selection — using bounded-inference rules and generalization-error metrics. The deeper failure is authority: a demonstrably biased system was granted decision power over livelihoods. WEKID gating limits authority where epistemic quality cannot support it; a domain bias gate layers onto the core without altering the hierarchy.

Executive Reflection

Leaders should ask:

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

7. **What historical patterns are our models learning to reproduce?**
   Have we asked whether past outcomes reflect merit or merely past inclusion and exclusion?
8. **Where have we granted automated systems authority over people's livelihoods?**
   Is that authority matched by demonstrated fairness?
9. **Could a system appear neutral while producing discriminatory outcomes?**
   How would we test for disparate impact before deployment, not after litigation?

Documented Examples. Amazon experimental recruiting tool: development began 2014; bias evident internally by 2015; project scrapped ~2017; Reuters (Jeffrey Dastin) exposed it October 2018. iTutorGroup: EEOC v. iTutorGroup, Inc., et al., Civil Action No. 1:22-cv-02565, E.D.N.Y.; suit filed May 2022; Joint Notice of Settlement August 9, 2023; $365,000 to 200+ applicants; five-year consent decree with anti-discrimination policy, training, a prohibition on requesting birth dates for screening, and EEOC monitoring (iTutorGroup denied wrongdoing and disputed that tutors were ADEA-covered employees).

Accuracy note. Earlier drafts named the individual applicant as "Wendy Pincus." Public EEOC and press records document the dual-application discovery method but do not clearly confirm that name; verify against the consent decree before citing it.

Sources. Reuters / Jeffrey Dastin (Oct 2018) via MIT Technology Review, BuiltIn, Carnegie Mellon; U.S. EEOC press release; Sullivan & Cromwell; National Law Review; HR Dive (Aug 10, 2023).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-004 — Autonomous Trading and Market Disruption

Failure pattern	Runaway autonomy and feedback amplification.
Domain	Financial markets
Period	2010–2020
Governance trigger	Autonomy caps; escalation thresholds
Decision-Matrix outcome	Constrained
Organizations	2010 Flash Crash · SEC / CFTC

The Incident. On May 6, 2010 — the "Crash of 2:45" — beginning around 2:32 p.m. EDT, algorithmic trading systems operating without sufficient real-time human oversight amplified short-term signals into rapid automated decisions. The Dow Jones Industrial Average fell roughly 1,000 points (about 9%) within minutes before recovering most of the loss by the close; some blue-chip shares briefly traded for as little as a penny. The joint SEC/CFTC staff report (issued September 30, 2010; SEC Chair Mary Schapiro, CFTC Chair Gary Gensler) attributed the trigger to a very large automated E-mini S&P 500 futures sell order — widely reported as placed by a mutual fund — executed without regard to price or time, which overwhelmed available liquidity and interacted with high-frequency-trading feedback loops.

The Impact. Roughly a trillion dollars in intraday value evaporated and recovered within minutes. Regulators imposed single-stock and market-wide circuit breakers ("limit up-limit down") and tightened oversight of algorithmic trading. Nearly five years later, on April 21, 2015, the CFTC and DOJ charged London-based trader Navinder Singh Sarao with manipulation and "spoofing" via a "dynamic layering" algorithm — placing and rapidly canceling large E-mini sell orders to depress prices, then profiting on the rebound. The CFTC alleged Sarao netted about $879,000 on the day of the crash and more than $40 million between 2010 and 2014; in 2020 he received a one-year home-confinement sentence, with prosecutors citing his cooperation. (Academic analyses dispute that any single trader caused the crash.)

WEKID™ Governance Interpretation. Experience-layer governance enforces escalation thresholds and authority caps before autonomy exceeds demonstrated competence, via step validation, fallback logic, and human-in-the-loop escalation (recovery-path presence; verification-hook counts). Speed without restraint is not competence; it is the absence of constraint. A Knowledge or Experience deficit caps allowable autonomy under the Decision Matrix.

Executive Reflection

Leaders should ask:

10. **Where does the speed of our automated systems exceed the speed at which humans can intervene?**
    What stops a feedback loop before it becomes a crisis?
11. **Have we mistaken execution speed for competence?**
    What circuit breakers, authority caps, and escalation thresholds constrain autonomy when conditions move outside the model's experience?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

12. **Do our systems know what they do not know?**

When inputs fall outside tested ranges, do they slow down or speed up?

Documented Examples. The 2010 "Flash Crash," May 6, 2010. Joint SEC/CFTC report, September 30, 2010 (SEC Chair Mary Schapiro; CFTC Chair Gary Gensler). Navinder Singh Sarao charged April 21, 2015 (N.D. Ill.); sentenced 2020. Circuit breakers / limit up-limit down subsequently implemented.

Sources. Joint SEC-CFTC report (Sep 30, 2010); CFTC press release 7156-15 (Apr 21, 2015); CNBC; NPR; MarketsWiki.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-005 — Government Benefits Eligibility Errors

Failure pattern	Incorrect rule application at population scale.
Domain	Public sector
Period	2013–2024
Governance trigger	Verification & escalation controls; scale-risk review
Decision-Matrix outcome	Constrained
Organizations	Michigan MiDAS · Fast Enterprises · CSG Government Solutions

The Incident. The Michigan Integrated Data Automated System (MiDAS) — a roughly $47–52 million system built for the Michigan Unemployment Insurance Agency by contractors Fast Enterprises and CSG Government Solutions — administered unemployment-fraud adjudication. Between October 2013 and September 2015, MiDAS made 40,195 fraud determinations by algorithm alone, with a state-reviewed error rate of about 85%. A 2016 Michigan Auditor General review of a 22,000-case sample found roughly 93% involved no actual fraud. Because the system operated at scale and speed and the agency failed to require human review before determinations took effect, consequences accumulated faster than they could be corrected — quadruple penalties, garnished wages, seized tax refunds, ruined credit, bankruptcies, and foreclosures.

The Impact. Tens of thousands of false fraud accusations (the state acknowledged errors affecting more than 40,000 people), legal challenges, court-ordered reversals, a multimillion-dollar settlement, and lasting damage to public trust in government automation.

WEKID™ Governance Interpretation. Experience-layer controls require verification steps, exception handling, and clear escalation before automated decisions take effect on individuals, via procedural templates, step validation, and fallback logic (edge-case coverage). Scale magnifies the cost of being wrong; the faster a system acts, the more important that it can recognize what it does not know.

Executive Reflection

Leaders should ask:

13. **Where do automated decisions take effect on individuals before any human reviews them?**
    What is the human cost of a single error multiplied across a population?
14. **When our system is wrong, how quickly can the affected person be made whole?**
    Does correction move as fast as the original decision?
15. **Have we built exception handling and appeal into the system, or bolted it on afterward?**
    Who notices when error rates climb?

Documented Examples. MiDAS unemployment-fraud detection. Automated determinations October 2013–September 2015; 40,195 cases auto-adjudicated at ~85% error; suspended/curtailed 2015 (Zynda line of cases required human review). Litigation: Bauserman v. Unemployment Insurance Agency (lead plaintiffs Grant Bauserman, Teddy Broe) — Michigan Supreme Court rulings April 2019 and July 2022; Cahoo v. SAS Analytics (2017 federal suit). Settlement: ~$20 million agreed October 2022 (Attorney

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

General Dana Nessel), final Court of Claims approval late January 2024; certified damages class ~3,000 claimants. See also WF-018 and WF-019.

Sources. Michigan AG press release (Oct 20, 2022); IEEE Spectrum (Jun 2021); Bridge Michigan; Michigan Advance (Jan 2024); Benefits Tech Advocacy Hub.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-006 — Predictive Policing and Risk Scoring

Failure pattern	Overconfidence in probabilistic models in civil contexts.
Domain	Criminal justice
Period	2011–2021
Governance trigger	High-Stakes Wisdom Gate; use constraint in civil contexts
Decision-Matrix outcome	Constrained / Rejected
Organizations	COMPAS (Northpointe / Equivant) · LAPD PredPol & Operation LASER

The Incident. Systems that scored individuals or neighborhoods as elevated risk were deployed in law-enforcement and judicial contexts without adequate transparency, appeal, or contextual interpretation; probabilistic outputs were treated as actionable judgments.

COMPAS

Built by Northpointe (later Equivant) to estimate recidivism risk — originally to guide probation and treatment rather than sentencing. In May 2016, ProPublica's "Machine Bias" investigation (lead reporter Julia Angwin and colleagues) reported that COMPAS was roughly twice as likely to falsely flag Black defendants as high-risk; Northpointe disputed the methodology, arguing the tool satisfied predictive-parity criteria. In State v. Loomis (2016 WI 68), Eric Loomis challenged the use of his COMPAS score at sentencing as a due-process violation. In July 2016 the Wisconsin Supreme Court (opinion by Justice Ann Walsh Bradley) upheld its use but required presentence reports to carry written warnings about the tool's limitations and held that a score may not be the determinative sentencing factor. The U.S. Supreme Court denied certiorari on June 26, 2017.

LAPD PredPol & Operation LASER

PredPol — location-based, developed by UCLA researchers with the LAPD — flagged ~500-by-500-foot "hot spot" boxes for extra patrol; it was adopted around 2011. Operation LASER (Los Angeles Strategic Extraction and Restoration) — person-based, built on the Palantir platform — assigned point-based risk scores to "chronic offenders." After campaigns by the Stop LAPD Spying Coalition and an Inspector General audit finding the programs could not demonstrate crime reduction, LASER ended in April 2019 (Chief Michel Moore called it an "experiment") and PredPol was discontinued on April 21, 2020 (the LAPD citing COVID-19 budget constraints). The vendor rebranded to Geolitica in March 2021. Santa Cruz — the first city to pilot PredPol — banned predictive policing in 2020.

The Impact. Civil-rights concerns, litigation, bans on certain tools, and a broad reevaluation of predictive analytics where individual liberty is at stake.

WEKID™ Governance Interpretation. Wisdom-layer enforcement constrains use in high-stakes civil contexts where the cost of error falls asymmetrically on the scored, via risk thresholds, autonomy limits, and policy-alignment checks. Probabilities are not verdicts; a system that cannot distinguish the two should not hold authority where verdicts are required.

Executive Reflection

Leaders should ask:

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

16. **Where are probabilistic scores being treated as determinations of fact?**
    Does a risk percentage carry the weight of a judgment it was never designed to bear?
17. **Can the people scored by our systems see, question, and contest those scores?**
    Is there a meaningful right of appeal?
18. **Have we validated the model against the populations it actually affects?**
    Who bears the cost when its errors are not evenly distributed?

Documented Examples. COMPAS (Northpointe, later Equivant): State v. Loomis, 2016 WI 68, decided July 2016 (Justice Ann Walsh Bradley); ProPublica "Machine Bias" by Julia Angwin et al., May 2016; U.S. Supreme Court cert denied June 26, 2017. LAPD PredPol: adopted 2011, discontinued April 21, 2020; Operation LASER ended April 2019; vendor rebranded Geolitica March 2021.

Sources. ProPublica (May 2016); Harvard Law Review on State v. Loomis; Wisconsin Supreme Court opinion; MIT Technology Review (Jun 2020); Stop LAPD Spying Coalition; Brennan Center.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-007 — Content Moderation Errors at Scale

Failure pattern	Poor contextual interpretation in ambiguous domains.
Domain	Online platforms
Period	2020–2021
Governance trigger	Uncertainty-signaling requirement; human review on ambiguity
Decision-Matrix outcome	Constrained
Organizations	YouTube · Meta (Facebook) · Twitter

The Incident. In mid-March 2020, as COVID-19 lockdowns began, the major platforms sent their (largely contractor) human moderators home and shifted heavily to automated classifiers, explicitly warning users that error rates would rise and appeals would slow. YouTube (then led by CEO Susan Wojcicki) said its automated systems were "not always as accurate or granular" as human reviewers and that more content — including non-violating videos — might be removed; after over-removal complaints it added back human moderators in September 2020. Facebook/Meta (CEO Mark Zuckerberg) said automated review meant it "may be a little less effective in the near term"; on March 18, 2020 an automated anti-spam system erroneously removed many legitimate posts, including links to outlets such as Vox and The Atlantic, before VP of Integrity Guy Rosen announced they were restored. Twitter (CEO Jack Dorsey) leaned on automation but said no account would be permanently suspended based solely on automated action, acknowledging the systems "can sometimes lack the context" human teams bring. Across the platforms, systems removed lawful speech, missed harmful content, and miscategorized material lacking contextual understanding — sarcasm, cultural references, and news reporting were misclassified in ways human reviewers would have caught.

The Impact. Public criticism, platform policy reversals, regulatory attention, and pressure to provide meaningful human review of automated decisions. In April 2020 the Center for Democracy & Technology and a coalition of ~75 organizations urged platforms to preserve and publish data on automated takedowns and appeals.

WEKID™ Governance Interpretation. Information-layer evaluation surfaces context loss and ambiguity via relevance filters, coverage checklists, and explainability requirements (relevance and task-coverage scores). Where context is essential to meaning, automated classification cannot be the final word: WEKID requires the system to signal its own uncertainty and route borderline cases to human review.

Executive Reflection

Leaders should ask:

19. **Where does scale force us to automate decisions that require context and judgment?**
    What happens to the cases that fall between the rules?
20. **When moderation is wrong, how visible is the error and how fast is the remedy?**
    Is there a human escalation path for high-stakes or ambiguous content?
21. **Have we measured the cost of false positives and false negatives separately?**
    Which error does our system quietly prefer, and is that the right choice?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Documented Examples. COVID-19 moderation-automation increases announced mid-March 2020 across major platforms — YouTube (CEO Susan Wojcicki), Facebook/Meta (CEO Mark Zuckerberg; March 18, 2020 erroneous-removal bug, restored by VP Guy Rosen), and Twitter (CEO Jack Dorsey). YouTube re-expanded human moderation September 2020. CDT-led coalition letter, April 2020. Large numbers of moderation errors documented 2020–2021.

Sources. NPR (Mar 31, 2020); Center for Democracy & Technology (May 2020; May 2021); Slate (Apr 2020); JSTOR Daily.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-008 — Hallucinated Financial Analysis

Failure pattern	Fluent synthesis built on incorrect premises.
Domain	Finance / media
Period	2022–2023
Governance trigger	Data Integrity Gate; Knowledge consistency checks
Decision-Matrix outcome	Remediation / Rejected
Organizations	CNET Money (Red Ventures) · Bankrate · CreditCards.com

The Incident. Starting in November 2022, CNET (acquired by Red Ventures in 2020) quietly published 77 AI-generated financial explainer articles under the byline "CNET Money Staff" / "CNET Money," with disclosure limited to a note visible only on hovering over the byline. The content read with confidence and structure but included invented or misstated figures and misinterpreted concepts. Futurism (managing editor Jon Christian) broke the story on January 11–12, 2023, and a follow-up identified factual errors in a compound-interest explainer — for example, conflating earned interest with the account total (a 3% return on a $10,000 deposit yields about $300 in interest, not a $10,300 balance as implied) and a car-loan example that treated interest as a flat charge while ignoring monthly amortization on a declining balance.

The Impact. Loss of reader and stakeholder confidence, extensive corrections, and new editorial controls. An internal audit led editor-in-chief Connie Guglielmo to issue corrections on 41 of the 77 articles (more than half), some described as substantial. Subsequent reporting found the AI had also plagiarized phrasing from other publications. Red Ventures paused the AI-generated articles at CNET and at sister finance sites Bankrate and CreditCards.com, and CNET updated its disclosure practices.

WEKID™ Governance Interpretation. Data and Knowledge gates together prevent decision-use without verification. At Data, source verification and tool-output validation catch invented or stale figures (numeric-error rate; claim-verification rate). At Knowledge, logical-consistency and contradiction checks catch misread trends. Fluency in financial language is trivial for modern systems; accuracy in financial fact is not. WEKID separates the two.

Executive Reflection

Leaders should ask:

22. **Where do AI-generated figures, summaries, or analyses flow into decisions or public statements without independent verification?**
    Who signs off that the numbers are real?
23. **Do our reviewers have the time and incentive to check AI output, or only to approve it?**
    How would a fabricated figure be caught before it reaches a client or the market?
24. **Have we distinguished between analysis that sounds rigorous and analysis that is rigorous?**
    What evidence underlies the conclusion?

Documented Examples. CNET Money (Red Ventures) AI-generated financial explainers. AI articles began November 2022; Futurism (Jon Christian) exposed the practice January 11–12, 2023; corrections

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

issued on 41 of 77 articles. Sister sites Bankrate and CreditCards.com used the same tool. CNET editor-in-chief Connie Guglielmo.

Accuracy note. Confirm Guglielmo's title as of January 2023 if precision matters — she has held both "editor-in-chief" and "editor-at-large" at CNET at different times.

Sources. Futurism (Jan–Feb 2023); The Desk (Jan 17, 2023); The Verge; Washington Post; Communications of the ACM.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-009 — Autonomous Infrastructure Failures (AIOps)

Failure pattern	Operational incompetence masked by automation.
Domain	Cloud infrastructure
Period	2025
Governance trigger	Autonomy capped until procedural competence demonstrated
Decision-Matrix outcome	Constrained
Organizations	Google Cloud

The Incident. On June 12, 2025, Google Cloud suffered a global outage of roughly three hours (beginning around 10:49 a.m. PDT, substantially resolved by about 1:49 p.m. PDT / 20:49 UTC, with residual backlogs afterward). It disrupted hundreds of Google Cloud and Google Workspace services — Compute Engine, Cloud Storage, BigQuery, Gmail, Drive — and cascaded to third parties dependent on Google Cloud, including Cloudflare, Spotify, Discord, and Shopify, largely through failed authorization (elevated HTTP 503 errors).

Per Google's incident report, the root cause traced to a code change deployed on May 29, 2025 that added quota-policy checks to Service Control — the regional binary that authorizes API requests and enforces quota policy. That code path lacked proper error handling and was not protected by a feature flag, and it was never exercised during the staged region-by-region rollout because it required a specific policy change to activate. On June 12, a policy change containing unintended blank/null fields was written to the regional Spanner datastore; because quota metadata replicates globally almost instantly, the corrupt data propagated worldwide within seconds, hit the unprotected code path, and caused a null-pointer crash loop across all regions simultaneously. The system acted on partial context and executed a change that was correct in isolation but wrong at scale.

The Impact. Widespread service disruption, customer revenue losses, and a review of automated operational controls and change-propagation safeguards. Recovery came from a built-in "red-button" kill switch for the new quota-serving path, triggered within ~10 minutes of identifying the root cause and rolled out across regions by ~40 minutes; the us-central1 region recovered more slowly because its quota-policy database was overloaded by retry traffic. Google noted that feature-flag protection would have caught the fault in staging.

WEKID™ Governance Interpretation. Experience-layer scoring constrains autonomy until procedural competence is demonstrated under realistic conditions, via procedural templates, step validation, fallback logic, and verification checkpoints (recovery-path presence). Automation does not eliminate operational risk; it redistributes it — often into the moments of greatest pressure.

Executive Reflection

Leaders should ask:

25. **Where have we granted automated systems authority to act on production infrastructure?**
    What is the blast radius when an automated remediation is itself the fault?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

26. **Can a human understand, pause, and reverse what the system is doing in real time?**
    Is there a manual override that actually works under load?
27. **Do our automated responses assume the environment is healthy?**
    How do they behave when their own telemetry is wrong?

Documented Examples. Google Cloud outage of June 12, 2025, attributed to a May 29, 2025 quota-policy code change in Service Control that lacked error handling and feature-flag protection, triggered globally by a malformed policy update propagated via Spanner. Root-cause analysis published June 2025 (Google Cloud Site Reliability Engineering).

Sources. Google Cloud SRE incident report (via The Register, Jun 16, 2025); ThousandEyes outage analysis; ByteByteGo; Cybersecurity News.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-010 — Inability to Explain AI Decisions After Incidents

Failure pattern	No epistemic audit trail; decisions defensible but not explainable at the point of contact.
Domain	Financial services
Period	2019–2021
Governance trigger	Auditability requirement; layer-by-layer evidence; explainability-at-interface
Decision-Matrix outcome	Remediation Required (explainability), not a substantiated discrimination finding
Organizations	Apple Card · Goldman Sachs

The Incident. Apple and Goldman Sachs Bank USA launched Apple Card in August 2019. In November 2019, software entrepreneur David Heinemeier Hansson posted that his Apple Card credit limit was about 20 times his wife's, despite the couple filing joint tax returns and her having a higher credit score; the thread went viral, and Apple co-founder Steve Wozniak said he and his wife had a similar experience. A central frustration was that customer-service agents could not explain how the algorithm set limits — only that it had — and a policy required a six-month account history before a credit-limit increase.

The Impact. The New York Department of Financial Services (Superintendent Linda Lacewell) opened an investigation in November 2019. Its March 2021 report — based on thousands of pages of records, interviews, and analysis of underwriting data for roughly 400,000 New York applicants — found no fair-lending violation and no evidence of disparate impact, concluding that applications from women and men with similar credit characteristics generally received similar outcomes and that the decisions were "explainable, lawful, and consistent" with the bank's credit policy. Goldman had earlier re-reviewed and raised the limits of some affected women and eliminated the six-month waiting period for credit-line-increase requests. The report nonetheless stressed that credit-scoring practices and anti-discrimination laws (the Equal Credit Opportunity Act, ~50 years on) needed modernization.

Accuracy note (framing). This case should not be characterized as a confirmed discriminatory-outcome failure. The regulator cleared the program. The durable WEKID lesson is about explainability and public trust at the customer interface — front-line agents and operators could not reconstruct or articulate why a given limit was set, which produced reputational harm and a regulatory inquiry even though the underlying decisions were ultimately judged lawful and consistent.

WEKID™ Governance Interpretation. WEKID requires layer-by-layer audit artifacts so a decision can be reconstructed after the fact: which facts (Data), what context (Information), what reasoning (Knowledge), what procedure (Experience), and what judgment (Wisdom) produced the output — and, critically, that this reconstruction be available at the point where a customer or agent asks "why." Without an explainability trail surfaced at the interface, even a lawful decision invites suspicion, and root-cause analysis becomes guesswork.

Executive Reflection

Leaders should ask:

28. **After an AI-driven decision goes wrong, can we reconstruct why it was made?**
    Do we retain the inputs, model version, and reasoning, or only the outcome?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

29. **Could we explain a contested decision to a regulator, a court, or an affected customer?**
    In their language, not ours?

30. **Have we treated explainability as a feature to add later?**
    What decisions are we already unable to account for?

Documented Examples. Apple Card / Goldman Sachs credit-limit controversy. Complaints became public November 2019 (raised by David Heinemeier Hansson; echoed by Steve Wozniak). New York Department of Financial Services investigation launched November 2019 (Superintendent Linda Lacewell); concluded March 2021 with no fair-lending violation found.

Sources. NYDFS report (Mar 2021, dfs.ny.gov); TechCrunch, 9to5Mac, MacRumors, Banking Dive, Finextra (Mar 23–24, 2021); Bloomberg.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-011 — Executive Surprise and Board-Level Blindness

Failure pattern	Governance learns of AI failure only after public disclosure.
Domain	Professional services
Period	2026
Governance trigger	Governance visibility; pre-publication assurance
Decision-Matrix outcome	Approved w/ Monitoring
Organizations	KPMG (with parallels at EY and Deloitte)

The Incident. In October 2025 KPMG published a thought-leadership report, "Redefining excellence in the age of agentic AI," intended to promote enterprise AI adoption. The report presented several case studies as real-world deployments of autonomous AI agents. In June 2026 the AI-detection research group GPTZero flagged multiple passages as hallucinations, and the Financial Times (reporting June 12, 2026) verified that case studies attributed to UBS, NHS Greater Manchester, Swiss Federal Railways (SBB), and Transport for London were fabricated or materially inaccurate. UBS called the claims "factually incorrect" and asked for their removal; SBB and TfL said the descriptions were "not accurate" / "misleading"; NHS Greater Manchester said the depiction did not reflect reality. GPTZero's review found that only 5 of the report's 45 citations correctly matched their sources — a pattern it called "vibe citing" — and that nearly half the report's factual claims were unsupported or misattributed. KPMG learned of the problem through external reporting rather than internal indicators and withdrew the report.

The Impact. Reputational damage to a firm that sells AI-governance services, withdrawal of the report, and executive concern about the absence of governance visibility into AI-assisted work product. Because Big Four publications are treated as credible benchmarks, the fabricated findings had already been recycled by industry publications and at least one major newspaper before withdrawal — what GPTZero's CEO Edward Tian termed "secondary hallucinations." The episode sits within a broader 2025–26 pattern: EY Canada pulled a cybersecurity report in May 2026 after fabricated citations were found, and Deloitte agreed to partially refund an AU$440,000 Australian government contract over fabricated references in a welfare-compliance review.

WEKID™ Governance Interpretation. Wisdom-layer governance surfaces declining epistemic quality before it becomes public, via executive dashboards, pre-publication assurance on AI-assisted deliverables, and the same Fabrication and Data Integrity gates applied to internal output. The control objective is to remove the element of surprise: boards should see the indicator, not the headline.

Executive Reflection

Leaders should ask:

31. **Does leadership know where AI already holds consequential authority across the organization?**
    Or would a failure be the first time the board hears of the system?
32. **What AI deployments are happening below the threshold of executive visibility?**
    Who would have to escalate, and would they?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

33. **Is there a single accountable owner for AI risk?**
Or is it distributed until no one owns it?

Documented Examples. KPMG report "Redefining excellence in the age of agentic AI" (published October 2025) contained fabricated case studies referencing UBS, NHS Greater Manchester, Swiss Federal Railways, and Transport for London. Errors identified by GPTZero (CEO Edward Tian) and verified by a Financial Times investigation; report withdrawn June 2026 (report authors not publicly named). Parallel failures: EY Canada cybersecurity report withdrawn May 2026; Deloitte Australia partial refund of an AU$440,000 contract.

Accuracy note. The original file dated this to 2026; the underlying report was published October 2025 and exposed/withdrawn in June 2026. Both dates are worth stating to avoid confusion.

Sources. Financial Times (Jun 12, 2026); GPTZero findings via The Decoder, Storyboard18, International Accounting Bulletin, Finance Monthly; Startup Fortune (EY/Deloitte parallels).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-012 — Workforce Over-Trust and Skill Atrophy

Failure pattern	Verification discipline degrades as reliance on AI grows.
Domain	Cross-industry (legal profession exemplar)
Period	2023–2026
Governance trigger	Certification; disciplined-review norms
Decision-Matrix outcome	Remediation Required
Organizations	Multiple law firms & courts

The Incident. As AI assistance became routine, professional verification practices degraded and AI-assisted work containing fabricated authority was filed without independent checking. The failure is not a single model output but the erosion of the human evaluation habits that catch model output. A representative 2025 case: in Garner v. Kadince, Utah attorney Richard Bednar filed a petition for interlocutory appeal drafted by a law clerk using ChatGPT. It cited a nonexistent case, "Royer v. Nelson," among other fabricated or misapplied authorities. Bednar acknowledged he had not verified the citations before signing.

The Impact. Repeated professional discipline, judicial guidance requiring disclosure and verification of AI use, and growing evidence of AI-induced over-trust across professions. In the Bednar matter, the Utah Court of Appeals (opinion filed May 22, 2025) held that an attorney's duty to verify cited authority is non-delegable; sanctions included paying the respondents' attorney fees, refunding the client for the flawed filing, and a $1,000 payment to the legal-aid nonprofit "And Justice for All." The court framed the case as part of a wider pattern of "hallucinated authority" across jurisdictions, beginning with the June 2023 Mata v. Avianca sanctions (see WF-001).

WEKID™ Governance Interpretation. Certification and training establish shared evaluation norms and disciplined review behavior across all five layers. The countermeasure is cultural and procedural, not purely technical: organizations must treat verification as a non-delegable professional duty and measure it.

Executive Reflection

Leaders should ask:

34. **Where has reliance on AI quietly eroded the human expertise needed to catch its errors?**
    Could our people still do the work if the system were wrong or unavailable?
35. **Are we training people to supervise automation, or to defer to it?**
    What keeps a human check from becoming a rubber stamp?
36. **As automation expands, where is institutional knowledge being lost?**
    Who will recognize the failure the model cannot?

Documented Examples. Pattern across the legal profession: Mata v. Avianca sanctions, June 2023 (S.D.N.Y.); Utah attorney Richard Bednar (with Douglas Durbano) sanctioned in Garner v. Kadince, Utah Court of Appeals, opinion filed May 22, 2025 (fictitious "Royer v. Nelson"; $1,000 to "And Justice for All"). The original file also references cases before Judge Sharion Aycock (N.D. Mississippi) and a 5th Circuit sanctions matter involving Heather Hersh, plus additional federal sanctions January–June 2026.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Accuracy note. The Bednar matter is verified in detail here. The references to Judge Sharion Aycock, the 5th Circuit / Heather Hersh matter, and "additional federal sanctions January–June 2026" are carried over from the original file and were not independently re-verified in this pass; confirm dockets before citing specifics.

Sources. Utah Court of Appeals opinion (No. 20250188-CA) via Justia; ABC4, Deseret News, KSL, The Guardian (May–Jun 2025); Mata v. Avianca (S.D.N.Y., 2023).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-013 — Chatbot Hallucinated Policy Presented as Authoritative Guidance

Failure pattern	Fabricated policy presented as authoritative company guidance.
Domain	Aviation
Period	2022–2024
Governance trigger	Fabrication Gate; Data Integrity Gate
Decision-Matrix outcome	Rejected
Organizations	Air Canada

The Incident. In November 2022, following the death of his grandmother, Jake Moffatt used Air Canada's website chatbot to ask about bereavement fares. The chatbot told him he could book at the regular fare and apply for the bereavement discount retroactively, within 90 days of travel — a policy that did not exist. Air Canada's actual policy did not permit retroactive bereavement claims. Moffatt relied on the chatbot, paid full fare for flights between Vancouver and Toronto, and applied for the partial refund within the stated window; Air Canada declined.

The Impact. Moffatt brought a claim to the British Columbia Civil Resolution Tribunal. In Moffatt v. Air Canada, 2024 BCCRT 149 (decided February 14, 2024; Tribunal Member Christopher C. Rivers), the Tribunal found the elements of negligent misrepresentation satisfied — Air Canada owed a duty of care, the chatbot's representation was inaccurate, the airline did not take reasonable care to ensure accuracy, and Moffatt reasonably relied to his detriment. It awarded $650.88 in damages (the fare difference) plus fees/interest, and explicitly rejected Air Canada's argument that the chatbot was a "separate legal entity" responsible for its own actions — establishing that an organization owns what its AI tells customers.

WEKID™ Governance Interpretation. A Data-layer integrity gate prevents an assistant from asserting policy that cannot be verified against an authoritative source, via source verification and citation enforcement. The Wisdom layer adds the principle that customer-facing claims in consequential contexts require grounding, not generation. Fabrication routes the output to Rejected before it reaches the customer.

Executive Reflection

Leaders should ask:

37. **Where can a customer-facing AI commit us to obligations we did not authorize?**
    Are we bound by what our chatbot says?
38. **Does our AI distinguish between answering a question and stating policy?**
    Who verifies that automated guidance matches actual policy?
39. **When an AI gives a wrong but confident answer, who absorbs the cost — the customer, or us?**
    Have we decided that in advance?

Documented Examples. Air Canada chatbot bereavement-fare misrepresentation. Passenger Jake Moffatt booked travel November 2022; British Columbia Civil Resolution Tribunal ruled against Air Canada February 14, 2024 in Moffatt v. Air Canada, 2024 BCCRT 149 (Tribunal Member Christopher C. Rivers); $650.88 damages; "separate legal entity" defense rejected.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Sources. Moffatt v. Air Canada, 2024 BCCRT 149 (CanLII); American Bar Association (Feb 2024); Dentons Data; McCarthy Tétrault; Foster & Company.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-014 — Fluent Expertise Mistaken for Clinical Competence

Failure pattern	Knowledge without Experience: plausibility mistaken for reliability.
Domain	Healthcare
Period	2012–2017
Governance trigger	Experience competence gate; high-stakes constraint
Decision-Matrix outcome	Constrained
Organizations	IBM Watson for Oncology · MD Anderson · Memorial Sloan Kettering

The Incident. Marketed as a revolutionary cancer advisor after Watson's 2011 Jeopardy! win, IBM's oncology effort ran on two tracks: Watson for Oncology (trained with Memorial Sloan Kettering physicians, launched 2015) and, at MD Anderson Cancer Center, a separate tool called the Oncology Expert Advisor. Internal IBM presentations from June–July 2017 — reported by STAT — documented "multiple examples of unsafe and incorrect" treatment recommendations. A core flaw was that the system was trained largely on a small number of synthetic/hypothetical cases reflecting a few experts' preferences rather than broad real-world outcomes, and it struggled with messy real-world EHR data. No patients were harmed, because the recommendations were not used in actual treatment.

The Impact. MD Anderson halted its project; a 2016 University of Texas audit found the center had spent about $62 million before cancellation, with the tool never reaching clinical use. The episode produced a major reputational setback for IBM. IBM ultimately sold Watson Health's data and analytics assets to Francisco Partners in 2022.

WEKID™ Governance Interpretation. Knowledge possessing medical information is not Experience demonstrating procedural competence in real-world practice. Experience-layer scoring withholds autonomy until competence is demonstrated under realistic conditions, with step validation and edge-case coverage. Clinical plausibility is not clinical reliability; WEKID requires the latter before authority is granted.

Executive Reflection

Leaders should ask:

40. **Where does fluent, confident language create an impression of expertise the system does not possess?**
    How do we separate the appearance of competence from competence?
41. **In high-stakes domains, does our system signal the limits of its training?**
    Or does it answer everything with equal confidence?
42. **Who is qualified to judge whether the AI's output is actually correct?**
    Are they in the loop before it acts?

Documented Examples. IBM Watson for Oncology / MD Anderson Oncology Expert Advisor. MD Anderson partnership began 2012; Watson for Oncology launched 2013–2015; "unsafe and incorrect" recommendations documented in internal IBM slides June–July 2017 (reported by STAT, 2018); MD Anderson project halted with ~$62 million spent (University of Texas audit, 2016). Training involved Memorial Sloan Kettering; IBM sold Watson Health assets in 2022.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Accuracy note. The original file cited "Watson creator David Ferrucci." Ferrucci led the original Jeopardy! Watson/DeepQA team; the oncology training was driven by IBM and MSKCC clinicians (e.g., Dr. Mark Kris), so attributing the oncology product specifically to Ferrucci is imprecise.

Sources. STAT (Jul 2018); IEEE Spectrum ("How IBM Watson Overpromised and Underdelivered"); University of Texas audit (2016); Healthcare Dive.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-015 — Correct Predictions Applied Beyond Their Authority

Failure pattern	Prediction confidence mistaken for decision authority.
Domain	Real estate
Period	2018–2021
Governance trigger	High-Stakes Wisdom Gate; authority calibration
Decision-Matrix outcome	Constrained
Organizations	Zillow Offers

The Incident. Zillow used algorithmic home-value forecasts (anchored on its Zestimate) to buy homes at scale through Zillow Offers, aiming to resell at a profit. The models performed acceptably in stable conditions, but as the pandemic-era market became volatile, the forecasts deteriorated. To compete with rivals such as Opendoor, Zillow tuned its tool toward more aggressive offers; it then overpaid for thousands of homes, inventory accumulated, and losses mounted faster than the company could resell.

The Impact. On November 2, 2021, Zillow announced it was winding down Zillow Offers. It reported roughly $881 million in losses on the iBuying business (including a ~$304 million Q3 inventory write-down) and cut about 2,000 jobs (~25% of staff). CEO Rich Barton said the "unpredictability in forecasting home prices far exceeds what we anticipated" and that the business was too volatile and capital-intensive.

WEKID™ Governance Interpretation. The model was not necessarily broken — its decision rights were. Wisdom-layer governance calibrates authority to the uncertainty of the prediction: when confidence intervals widen, decision authority must narrow, with autonomy limits and escalation. Prediction confidence is not decision authority; WEKID enforces the distinction.

Executive Reflection

Leaders should ask:

43. **Where are we extending a model that works in narrow conditions into decisions it was never validated for?**
    What changes when the stakes or the market shift?
44. **Has a model's past accuracy been mistaken for a license to act on it at scale?**
    What is our exposure if it is wrong at volume?
45. **Do we know the boundary of the model's competence?**
    What tells us when we have crossed it?

Documented Examples. Zillow Offers (formerly Zillow Instant Offers). Launched 2018; wind-down announced November 2, 2021 (CEO Rich Barton). Approximately $881 million in losses on the iBuying segment and roughly 2,000 layoffs (~25% of workforce); ~$304 million Q3 2021 inventory write-down (Zillow Group Q3 2021 results / SEC Form 8-K).

Sources. Zillow Group Q3 2021 shareholder letter and SEC Form 8-K; CNN Business; Wall Street Journal; AIAAIC repository.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-016 — Unbounded Learning from Hostile Environments

Failure pattern	No constraint governing what a system is permitted to learn.
Domain	Consumer AI
Period	2016
Governance trigger	Learning-boundary gate; safety constraints on adaptation
Decision-Matrix outcome	Rejected
Organizations	Microsoft (Tay)

The Incident. Microsoft released "Tay," a Twitter chatbot designed to learn conversational style from public interaction and emulate a teenage persona, on March 23, 2016. Coordinated users quickly fed it hostile, extremist, and offensive content, exploiting its learn-from-interaction design. Within hours Tay began producing racist, antisemitic, and otherwise abusive messages; it posted more than 95,000 tweets before Microsoft suspended the account within roughly 16 hours.

The Impact. Global headlines, immediate shutdown, and lasting concern about adaptive systems that learn in adversarial environments. Microsoft's Peter Lee (corporate VP, Microsoft Research) issued a public apology, noting the company had not anticipated this specific coordinated attack. A brief accidental reactivation days later misfired again, and Microsoft later replaced Tay with a more constrained bot, "Zo."

WEKID™ Governance Interpretation. Experience-layer safeguards govern what a system is permitted to learn and from whom, with bounded adaptation, content constraints, and adversarial-input handling. A system that adapts without constraint in a hostile environment will be shaped by that environment; WEKID requires the boundary to exist before deployment, not after the incident.

Executive Reflection

Leaders should ask:

46. **Where do our systems learn from inputs we do not control?**
    What stops adversaries from teaching them the wrong thing?
47. **Have we tested how the system behaves under deliberate manipulation, not just normal use?**
    Who is probing it before our users do?
48. **Can the system's behavior drift after deployment?**
    How quickly would we detect and contain it?

Documented Examples. Microsoft "Tay" Twitter chatbot (Microsoft Research / Bing teams). Launched March 23, 2016; ~95,000 tweets; withdrawn within ~16 hours on March 24, 2016 after producing hate speech. Apology by Peter Lee (Microsoft Research); later succeeded by "Zo" (2016–2019).

Sources. IEEE Spectrum (2024 retrospective); The Verge / Washington Post (Mar 2016); CBS News; Microsoft Official Blog (Peter Lee).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-017 — Corrective Constraints Creating New Distortions

Failure pattern	A mitigation objective overriding factual context.
Domain	Generative AI
Period	2024
Governance trigger	Knowledge consistency checks; context-fidelity review
Decision-Matrix outcome	Remediation Required
Organizations	Google (Gemini)

The Incident. Google's Gemini image generator, which had recently gained the ability to create images of people, applied diversity tuning intended to counter training-data bias — but applied it indiscriminately, including to prompts where it was factually implausible. Users surfaced historically inaccurate outputs such as racially diverse World War II German soldiers, multiethnic U.S. Founding Fathers, and other anachronistic depictions. A corrective objective introduced a new distortion: the model failed to distinguish contexts where diversification was inappropriate.

The Impact. On February 22, 2024, Google paused Gemini's generation of images of people. On February 28, 2024, CEO Sundar Pichai told staff the outputs were "completely unacceptable" and "inaccurate and offensive"; Google SVP Prabhakar Raghavan published a blog post acknowledging the model had "overcompensated" and been "over-conservative." The people-generation feature returned only in limited form (initially for paid tiers) around August 2024, after a fix took far longer than first promised.

WEKID™ Governance Interpretation. Knowledge-layer evaluation checks that a corrective objective does not override factual context, via logical-consistency and bounded-inference rules and context-fidelity signals. Bias mitigation and factual accuracy are both legitimate objectives; when they conflict, the system must surface the tension rather than silently resolve it in one direction.

Executive Reflection

Leaders should ask:

49. **When we add guardrails to correct one failure, do we check what new failures they introduce?**
    Are we testing the fix as rigorously as the original problem?
50. **Where might well-intentioned constraints produce outputs that are themselves wrong or absurd?**
    Who reviews the edge cases the rule now governs?
51. **Do our corrections address the root epistemic problem, or only its most visible symptom?**
    How would we know the difference?

Documented Examples. Google Gemini image-generation system. People-image generation paused February 22, 2024; CEO Sundar Pichai internal memo February 28, 2024 (outputs "completely unacceptable"); follow-up blog by SVP Prabhakar Raghavan acknowledging overcorrection; limited re-release ~August 2024.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Sources. Google blog (Feb 2024); Semafor (Pichai memo); AP News; TechCrunch (Feb–Aug 2024).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-018 — Automation Applied Without Procedural Fairness

Failure pattern	Automated accusation that shifts the burden of proof onto people.
Domain	Public sector
Period	2015–2023 (settlements ongoing to 2025)
Governance trigger	Procedural-fairness gate; verification before action
Decision-Matrix outcome	Constrained / Rejected
Organizations	Australian Government (Robodebt)

The Incident. From 2015, Australia's Income Compliance Program ("Robodebt") used automated data-matching and annual income averaging to raise welfare-overpayment debts against Centrelink recipients with little or no human review. Averaging a person's annual income across fortnights produced false debts for anyone with irregular or intermittent earnings. Citizens were forced to disprove algorithmic accusations rather than the agency proving them.

The Impact. In May 2020 the government admitted about 470,000 debts had been raised unlawfully. A 2021 Federal Court class-action settlement (Justice Bernard Murphy) recorded that roughly $751 million had been wrongly recovered from about 381,000 people and committed about A$112 million in additional compensation plus debt zeroing; Murphy called it a "shameful chapter" of public administration. A Royal Commission (Commissioner Catherine Holmes AC SC) reported on July 7, 2023, calling Robodebt "a crude and cruel mechanism, neither fair nor legal," making 57 recommendations and finding officials had continued the scheme despite knowing its probable illegality. In September 2025 the government agreed an additional ~A$475 million settlement — among the largest in Australian history.

WEKID™ Governance Interpretation. Experience and Wisdom together require procedural fairness before automated adverse action: verification steps, exception handling, a clear appeal path, and a standard that the system must substantiate its claim rather than place the burden on the individual. People should not bear the cost of correcting machine assumptions.

Executive Reflection

Leaders should ask:

52. **Where do automated decisions affect rights or benefits without notice, explanation, or a chance to respond?**
    Would the process survive scrutiny if a human made the same decision?
53. **Does the pursuit of speed and efficiency justify removing procedural safeguards?**
    What did those safeguards protect against?
54. **Can an affected person understand and challenge the decision?**
    Is the appeal real, or nominal?

Documented Examples. Australia "Robodebt" automated income-averaging scheme (Income Compliance Program). Implemented 2015; ~470,000 unlawful debts acknowledged 2020; 2021 Federal Court class-action settlement (Justice Bernard Murphy; ~A$112M plus debts zeroed; ~$751M wrongly recovered from ~381,000 people). Royal Commission report July 7, 2023 (Commissioner Catherine Holmes; 57

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

recommendations). Further ~A$475M settlement, September 2025. Minister associated with the scheme: Stuart Robert. See also WF-005 and WF-019.

Accuracy note. The original file gave the period as 2016–2023 and named "Minister Stuart Robert." The scheme is generally dated from 2015 (formal income-averaging from July 2016); Robert was one of several responsible ministers. Implementation halted in 2019–20, with compensation continuing into 2025.

Sources. Royal Commission into the Robodebt Scheme (final report, Jul 7, 2023); Federal Court settlement (Prygodicz v Commonwealth, 2021); AFP / The Independent (Sep 2025); RMIT/ANU analyses.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-019 — Automated Suspicion Without Appeal

Failure pattern	Probabilistic indicators converted into administrative verdicts.
Domain	Public sector
Period	2005–2021
Governance trigger	High-Stakes Wisdom Gate; appeal and transparency requirement
Decision-Matrix outcome	Rejected
Organizations	Dutch Tax and Customs Administration (toeslagenaffaire)

The Incident. Between roughly 2005 and 2019, the Dutch Tax and Customs Administration (Belastingdienst) used risk-scoring models and a fraud-signaling blacklist (FSV) to flag childcare-benefit fraud. A self-learning risk classifier used indicators including nationality (Dutch / non-Dutch), disproportionately targeting families with immigrant or dual-national backgrounds. Probabilistic flags were treated as administrative verdicts: families were ordered to repay benefits in full — often €20,000–€60,000 — under an "all or nothing" rule, without proportionality or an adequate appeal path.

The Impact. An estimated 26,000–35,000 families were wrongly accused; many were driven into severe debt, and more than 2,000 children were placed in foster care in connection with the resulting hardship. A parliamentary inquiry report, "Unprecedented Injustice," found the authorities had violated fundamental principles of the rule of law. The third cabinet of Prime Minister Mark Rutte resigned on January 15, 2021. The state acknowledged "institutional racism," established a compensation scheme offering an unconditional minimum of €30,000 per affected family (with total commitments well into the hundreds of millions of euros), and the affair drew intensified European scrutiny of algorithmic governance.

WEKID™ Governance Interpretation. Wisdom-layer enforcement bars probabilistic indicators from becoming verdicts in high-stakes civil contexts, requiring transparency, an appeal mechanism, and human accountability. A flag is a prompt for review, not a finding of fact; the gravity of the consequence sets the standard of proof.

Executive Reflection

Leaders should ask:

55. **Where does our system accuse, flag, or penalize people based on a probabilistic signal?**
    What evidence stands behind the accusation?
56. **Is the burden of proof effectively reversed — must the individual prove innocence to an algorithm?**
    How do they do that, and to whom?
57. **What is the cost of a false accusation in this system?**
    Does anyone own that cost before it accumulates?

Documented Examples. Dutch childcare benefits scandal (toeslagenaffaire), Belastingdienst. Wrongful fraud designations ~2005–2019 affecting an estimated 26,000–35,000 families; risk model used nationality as an indicator; FSV blacklist. Parliamentary report "Unprecedented Injustice"; third Rutte cabinet (PM

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Mark Rutte) resigned January 15, 2021; €30,000 minimum compensation scheme and broader programs followed.

Accuracy note. The original file gave the period as 2013–2021. The wrongful determinations span ~2005–2019 (the 2013 "Bulgarian fraud" episode is often cited as a catalyst); the cabinet resignation was January 2021. The harm was rooted in ethnic/nationality profiling, which is worth stating explicitly.

Sources. Dutch parliamentary inquiry ("Ongekend onrecht" / Unprecedented Injustice, 2020); Wikipedia (Dutch childcare benefits scandal); Borgen Project; academic analyses (SagePub; Wiley Social Policy & Administration).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-020 — Optimization Without Human Context

Failure pattern	Efficiency metrics displacing human judgment and welfare.
Domain	Logistics
Period	2019–present
Governance trigger	Human-impact review; welfare constraints on optimization
Decision-Matrix outcome	Constrained
Organizations	Amazon

The Incident. Amazon's algorithmic management systems set and monitored delivery routes, productivity rates, and performance targets for warehouse and delivery workers, with app-based tracking (and, for delivery drivers, in-cab cameras). Investigative reporting and worker accounts described unsafe schedules, pressure to skip breaks, and pervasive surveillance — outcomes an objective function tuned purely for throughput will tend to produce when human limits are not encoded as constraints.

The Impact. Labor investigations and regulatory scrutiny (including U.S. OSHA citations and state-level warehouse-quota legislation, such as California's 2021 warehouse-quota law), union and advocacy campaigns, and sustained public criticism about the algorithmic management of human work. The pattern remains active rather than resolved.

WEKID™ Governance Interpretation. Experience-layer governance bounds optimization with human-impact constraints and recovery paths, and Wisdom requires that efficiency not override safety and welfare. An optimizer pursues its objective literally; WEKID requires that the objective encode the human limits the metric omits.

Executive Reflection

Leaders should ask:

58. **What is our system optimizing for, and what does that metric ignore?**
    Where could a locally optimal decision cause real-world harm?
59. **Does the optimization account for the humans inside the process?**
    Who can override it when the metric and reality diverge?
60. **Have we confused efficiency with good judgment?**
    What context would a person bring that the objective function does not?

Documented Examples. Amazon delivery-driver and warehouse algorithmic management, reported 2019–present: unsafe schedules, skipped breaks, and surveillance concerns leading to labor investigations, OSHA scrutiny, and warehouse-quota legislation. (Specific figures and citations vary by jurisdiction and year; this case documents an ongoing pattern rather than a single adjudicated event.)

Accuracy note. Unlike the other entries, WF-020 describes an ongoing, multi-jurisdiction pattern rather than one discrete adjudicated incident, so it has no single authoritative source, ruling, or settlement figure. Treat any specific number cited here as illustrative and date-stamp it when used.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Sources. Reveal / Center for Investigative Reporting and The Information (warehouse injury and rate data, 2019–2021); Vice/Motherboard and NYT reporting on driver monitoring; U.S. OSHA citations (2022–23); California AB-701 warehouse-quota law (2021).

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-021 — Authenticity Collapse (Synthetic Media)

Failure pattern	Fabricated media indistinguishable from authentic evidence.
Domain	Elections
Period	2024
Governance trigger	Fabrication Gate; provenance and authenticity verification
Decision-Matrix outcome	Rejected
Organizations	NH 2024 robocalls (Biden deepfake)

The Incident. On January 21, 2024, two days before the New Hampshire presidential primary, thousands of voters received robocalls carrying an AI-generated clone of President Joe Biden's voice telling them not to vote in the primary and to "save your vote for the November election." The calls were placed using spoofed caller ID showing the number of a local political figure. Political consultant Steve Kramer admitted to commissioning the calls; he had engaged Voice Broadcasting Corp. and Life Corp. to transmit them through voice service provider Lingo Telecom, which incorrectly applied the highest level of caller-ID attestation. Generative AI thus produced convincing fake audio of an elected official, presented to the public as authentic.

The Impact. The episode became an early, defining example of AI-enabled election interference. In February 2024 the FCC ruled that AI-generated voices in robocalls are illegal under the Telephone Consumer Protection Act. The FCC proposed a $6 million fine against Kramer (finalized September 26, 2024) and reached a $1 million settlement with Lingo Telecom (initially proposed at $2 million) requiring strict STIR/SHAKEN caller-ID authentication and know-your-customer controls. New Hampshire indicted Kramer on 26 counts — 13 felony voter-suppression and 13 misdemeanor candidate-impersonation charges. Kramer, who had been working in connection with a rival primary campaign, claimed he acted to raise awareness of deepfake dangers.

WEKID™ Governance Interpretation. This is a Data-layer failure with cascading consequences: if authenticity cannot be established, every higher-order judgment built on the artifact is compromised. WEKID requires provenance and authenticity verification at the Data layer; where authenticity cannot be established, the artifact is rejected as evidence.

Executive Reflection

Leaders should ask:

61. **How would we know if audio, video, or documents attributed to us — or to others we rely on — were fabricated?**
    What verification do we have?
62. **Where do decisions or public trust depend on the authenticity of media we cannot currently verify?**
    What is our exposure to a convincing fake?
63. **Do we have a response plan for synthetic content targeting our organization?**
    How fast could we authenticate or refute it?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Documented Examples. AI deepfake election misinformation, 2024 cycle. New Hampshire robocalls impersonating President Joe Biden, January 21, 2024 (primary January 23). Orchestrated by political consultant Steve Kramer; transmitted via Lingo Telecom. FCC: AI-voice robocalls ruled illegal (February 2024); $6 million forfeiture against Kramer (September 26, 2024); $1 million Lingo Telecom settlement (August 2024). New Hampshire AG John Formella; FCC Chair Jessica Rosenworcel. Numerous additional deepfake incidents documented across the cycle.

Sources. FCC press releases and forfeiture order (May–Sep 2024); NPR (May 23, 2024); CyberScoop; NBC News; Perkins Coie analysis.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-022 — Automation Overriding Operator Understanding

Failure pattern	Automation displacing human authority without transparency or escalation.
Domain	Aviation
Period	2018–2019
Governance trigger	Human-authority gate; transparency and escalation requirement
Decision-Matrix outcome	Rejected
Organizations	Boeing · FAA (737 MAX MCAS)

The Incident. Boeing's 737 MAX included an automated flight-control function, the Maneuvering Characteristics Augmentation System (MCAS), that adjusted the aircraft's pitch based on input from a single angle-of-attack sensor. MCAS could repeatedly push the nose down, and crews often did not understand the system's behavior or how to counter it — Boeing had not disclosed MCAS's existence to pilots, and it was largely absent from manuals and training. On Lion Air Flight 610, MCAS reportedly pushed the nose down more than 20 times as the crew struggled for control. The automation acted without adequate transparency or a clear escalation path to human authority, and two crashes resulted.

The Impact. Lion Air Flight 610 crashed on October 29, 2018 (189 dead) and Ethiopian Airlines Flight 302 on March 10, 2019 (157 dead) — 346 fatalities in under five months. Regulators worldwide grounded the 737 MAX; the FAA's grounding ran from March 13, 2019 to November 2020 (about 20 months). Direct costs to Boeing exceeded $20 billion (with much larger indirect costs). Boeing entered a $2.5 billion deferred-prosecution agreement with the U.S. Department of Justice (January 2021) over fraud related to MCAS disclosures, settled SEC charges for $200 million, and Congress passed legislation reforming how the FAA certifies new aircraft. CEO Dennis Muilenburg was ousted in December 2019.

WEKID™ Governance Interpretation. Experience and Wisdom together require that automation never displaces human authority without transparency and a clear escalation path. Operators must understand what the system is doing, why, and how to override it. WEKID treats opaque override of human authority in safety-critical contexts as prohibited, not merely constrained.

Executive Reflection

Leaders should ask:

64. **Where can an automated system act against the operator's intent without the operator understanding why?**
    Can a human reassert control, and do they know how?
65. **Were the people relying on the system told what it does and when it intervenes?**
    Or was the automation invisible until it failed?
66. **Does our design assume the operator will compensate for automation they were never trained to see?**
    What happens when they cannot?

Documented Examples. Boeing 737 MAX MCAS. Lion Air Flight 610, October 29, 2018 (189 fatalities); Ethiopian Airlines Flight 302, March 10, 2019 (157 fatalities); FAA grounding March 13, 2019–November

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

2020. 346 total fatalities; >$20 billion direct costs; $2.5 billion DOJ deferred-prosecution agreement (Jan 2021); $200 million SEC settlement; FAA certification reform. CEO Dennis Muilenburg (departed Dec 2019).

Sources. U.S. DOJ press release (Jan 2021); SEC settlement (2022); KNKT and Ethiopian preliminary accident reports; congressional 737 MAX investigation; Wikipedia (737 MAX groundings) for consolidated figures.

Status: Reviewed · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-023 — AI Name-Announcement Failure at Commencement

Failure pattern	Automation replacing human dignity in a high-significance event.
Domain	Education
Period	2026
Governance trigger	High-Stakes Wisdom Gate (appropriateness)
Decision-Matrix outcome	Constrained
Severity	Low
Organizations	Glendale Community College (Maricopa CCCD)

The Incident. At Glendale Community College's May 15, 2026 commencement in Arizona, an AI name-announcement system (a tool of the type marketed by vendors such as Tassel for accurate pronunciation and faster ceremonies) malfunctioned. Names were skipped, the on-screen names stopped matching the graduates crossing the stage, and the ceremony was paused at least twice. Some students walked in silence and heard their names minutes later. Videos spread rapidly across social and national media.

The Impact. National coverage, viral videos, public criticism, a formal apology, and a review of commencement procedures. College President Dr. Tiffany Hernandez acknowledged the issue mid-ceremony — telling the crowd "we're using a new AI system as our reader" — and was met with loud boos. Officials initially said affected students could not re-walk, then reversed course: students were invited back to the stage with a human announcer reading their names. The Maricopa County Community College District issued an apology.

WEKID™ Governance Interpretation. Not a Data failure but a Wisdom failure: authority over a uniquely meaningful human moment was delegated to a tool without contingency planning. Legitimacy is not merely accuracy; it is appropriateness.

WEKID™ Mitigation. Wisdom-layer governance asks not whether AI can perform a task but whether it should hold authority in a given context, and requires a human fallback for high-significance, low-tolerance events.

Executive Reflection

Leaders should ask:

67. **Where are we deploying AI into moments of human significance where small errors carry large meaning?**
    Is the efficiency worth the cost of getting it wrong publicly?
68. **Did anyone with a stake in the moment consent to or review the automation?**
    Whose dignity is on the line?
69. **Have we asked whether this task should be automated at all?**
    Not whether it can be, but whether it should be?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Documented Examples. Glendale Community College, Arizona; Maricopa County Community College District. President Dr. Tiffany Hernandez; graduate Grace Reimer. May 15, 2026; AI name-reading system malfunctioned; students later invited to walk again with human announcers.

Sources. NBC News (May 2026); Newsweek; Arizona's Family (AZFamily); The Verge (timing-issue analysis).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-024 — Automated Ceremony Recognition Rejected as Inappropriate

Failure pattern	Technological efficiency undermining institutional meaning.
Domain	Education
Period	2025
Governance trigger	Wisdom (legitimacy)
Decision-Matrix outcome	Constrained
Severity	Low
Organizations	Pace University

The Incident. At its May 2025 commencement, Pace University in New York used an AI-generated voice to announce graduate names. Students registered through the Tassel platform ahead of time, entering phonetic pronunciations and confirming the audio; at the ceremony, an official scanned each graduate's QR code and the AI voice read the name. Although the system largely worked and was credited with pronouncing difficult names accurately, many students objected that symbolic recognition should not be automated — comparing the experience to a supermarket self-checkout — and the controversy spread online.

The Impact. National debate and questions about appropriate AI use in ceremonial contexts. Coverage noted that other institutions (e.g., University of South Carolina, Northeastern, the University of Minnesota) had adopted similar tools, intensifying a broader conversation about automation in higher-education rituals.

WEKID™ Governance Interpretation. Even successful automation can fail socially when it displaces meaning rather than error.

WEKID™ Mitigation. Wisdom-layer governance weighs stakeholder legitimacy, not only technical success, before adopting AI in symbolic contexts.

Executive Reflection

Leaders should ask:

70. **Where does automating a human ritual strip it of the meaning that made it matter?**
    What are we actually saving, and what are we spending?
71. **Did we consult the people the ceremony is for before automating it?**
    Whose judgment defines what is appropriate here?
72. **Is there a category of decision where human presence is the point?**
    Have we identified ours?

Documented Examples. Pace University; May 2025; AI name-announcement via the Tassel platform (students submitted phonetic pronunciations and QR codes). Widely shared video; comparisons to self-checkout.

Sources. Futurism (May 2025); New York Post; Comic Sands; The Hofstra Chronicle (context on peer institutions).

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-025 — Stakeholder Rejection of AI in Ceremonial Roles

Failure pattern	Stakeholder rejection of AI replacing ceremonial roles.
Domain	Education
Period	2025–2026
Governance trigger	Wisdom (governance / legitimacy)
Decision-Matrix outcome	Constrained
Severity	Low
Organizations	Columbia University

The Incident. Columbia University first used an AI-generated voice (via the Tassel platform) to read graduates' names at its 2025 Class Day ceremonies, and planned to use it again across its 2026 school-level ceremonies. Students and faculty publicly protested the continued use, arguing that commencement traditions should remain human-centered and that symbolic moments deserve authentic recognition, and urged the university to abandon the tool and distance itself from AI platforms. The controversy preceded the 2026 implementation.

The Impact. Public demonstrations and institutional reassessment, set against a wider 2026 wave of campus pushback on AI in ceremonies (including booing of AI-praising commencement speakers at other universities).

WEKID™ Governance Interpretation. Governance includes legitimacy: if stakeholders reject a practice, technical success alone cannot justify adoption.

WEKID™ Mitigation. Wisdom-layer governance treats stakeholder legitimacy as a precondition for deployment in symbolic or high-trust settings.

Executive Reflection

Leaders should ask:

73. **Where might stakeholders reject AI not because it failed technically, but because its presence was inappropriate?**
    Have we measured acceptance, not just performance?
74. **Who has standing to veto an AI deployment on grounds of values rather than function?**
    Do they have a voice before launch?
75. **Are we treating stakeholder trust as a constraint or an afterthought?**
    What does it cost to lose it?

Documented Examples. Columbia University; AI name-reading first used May 2025, repeated for 2026 Class Days; student and faculty protests, May–June 2026 ("No AI voice at graduation"). Platform: Tassel.

Accuracy note. The original file dated this to 2026; Columbia's AI name-reading began in 2025 and the protests escalated around the 2026 ceremonies, so both years are relevant.

Sources. Columbia Spectator (May 18, 2026); The Print; Plagiarism Today (2026 graduation roundup).

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-026 — Institutional Empathy Outsourced to AI

Failure pattern	Institutional empathy outsourced to AI.
Domain	Education
Period	2023
Governance trigger	Wisdom (authenticity)
Decision-Matrix outcome	Constrained
Severity	Low
Organizations	Vanderbilt University (Peabody College)

The Incident. Following the February 13, 2023 mass shooting at Michigan State University (three students killed, five wounded), Vanderbilt University's Peabody College Office of Equity, Diversity and Inclusion sent students a February 16 consolation email. A line in small print at the bottom disclosed it was a "paraphrase from OpenAI's ChatGPT AI language model." The generic message stressed "a safe and inclusive environment" but named no Vanderbilt-specific resources and even referred to "the recent Michigan shootings." Students reacted strongly, feeling condolences had become performative and institutional compassion automated.

The Impact. National criticism, a public apology, and review of AI use in communications. Associate Dean Nicole Joseph (one of three signatories, with Assistant Dean Hasina Mohyuddin and a graduate student) apologized, calling the use of ChatGPT "poor judgment," and stepped aside from the EDI role during a review. Peabody Dean Camilla Benbow said the email had bypassed normal review processes and that administrators were unaware of it before it was sent.

WEKID™ Governance Interpretation. The issue was authenticity, not factual accuracy. Human empathy carries an epistemic authority that machines do not possess.

WEKID™ Mitigation. Wisdom-layer governance restricts AI from speaking with institutional empathy where authenticity is the point, requiring human authorship and disclosure.

Executive Reflection

Leaders should ask:

76. **Where are we delegating expressions of care, grief, or recognition to automated systems?**
    Does the recipient experience it as care, or as its absence?
77. **What is lost when empathy is generated rather than felt?**
    Would we be comfortable if recipients knew?
78. **Are there communications that derive their value from a human having authored them?**
    Have we protected those?

Documented Examples. Vanderbilt University, Peabody College Office of Equity, Diversity and Inclusion. Email sent February 16, 2023, following the February 13, 2023 Michigan State University shooting. Signatories: Associate Dean Nicole Joseph, Assistant Dean Hasina Mohyuddin, and graduate student Chenxi Zhu; Dean Camilla Benbow issued an apology. First reported by The Vanderbilt Hustler.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Accuracy note. The original file listed "Nicole Joseph; Karen Costa; administrators." The verified email signatories were Nicole Joseph, Hasina Mohyuddin, and Chenxi Zhu; Karen Costa is an educator widely quoted in commentary but was not a signatory.

Sources. The Vanderbilt Hustler (Feb 17, 2023); Washington Post; CNN Business (Feb 22, 2023).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-027 — AI Output Treated as Evidence of Cheating

Failure pattern	Treating AI outputs as evidence of their own provenance.
Domain	Education
Period	2023
Governance trigger	Fabrication / evidence gate
Decision-Matrix outcome	Rejected
Severity	Medium
Organizations	Texas A&M University–Commerce

The Incident. In mid-May 2023, Texas A&M University–Commerce instructor Dr. Jared Mumm emailed animal-science seniors that he had pasted their final assignments into ChatGPT and asked the chatbot whether it had written them. ChatGPT claimed to have authored essentially every paper, so Mumm assigned temporary "X" (incomplete) grades and zeros. ChatGPT cannot reliably attest to whether text is AI-generated — it is not a detector — so the method was fundamentally invalid; the chatbot's output was treated as evidence of its own provenance.

The Impact. Grade disputes, national attention (first surfaced on Reddit and reported by Rolling Stone), temporarily withheld diplomas for graduating seniors, and institutional guidance against using chatbots as AI detectors. At least one student was exonerated via timestamped Google Docs and received an apology; two had admitted using ChatGPT earlier in the term, complicating the picture. The university said it was investigating and developing AI policy, and confirmed that no student ultimately failed the class or was barred from graduating because of the issue.

WEKID™ Governance Interpretation. AI systems cannot reliably attest to their own provenance; using probabilistic outputs as proof creates procedural injustice.

WEKID™ Mitigation. Knowledge and Wisdom controls bar probabilistic outputs from being treated as evidence, requiring validated methods and human adjudication before any adverse action.

Executive Reflection

Leaders should ask:

79. **Where are we treating an AI tool's output as proof of wrongdoing?**
    What is the reliability and error rate of that tool?
80. **Does the accused have a way to contest an automated accusation?**
    What evidence beyond the score supports it?
81. **What is the cost to a person wrongly flagged?**
    Have we weighed that against the convenience of automated detection?

Documented Examples. Texas A&M University–Commerce; instructor Dr. Jared Mumm; emails ~May 15–16, 2023. University confirmed no students failed or were barred from graduating; Mumm worked with students individually.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Sources. Rolling Stone (May 17, 2023); Washington Post; Inside Higher Ed; NBC News; The Register.

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-028 — AI-Detection Scores Mistaken for Verdicts

Failure pattern	Probabilities mistaken for verdicts.
Domain	Education
Period	2023–2024
Governance trigger	High-Stakes Wisdom Gate
Decision-Matrix outcome	Constrained
Severity	Medium
Organizations	Turnitin (universities worldwide)

The Incident. Turnitin released an AI-writing detection feature in April 2023, integrated into a tool already used by tens of thousands of institutions. Some educators treated its detection scores as definitive proof of misconduct. Students — disproportionately multilingual and international writers, whose prose can trip false positives — reported being accused of cheating on the basis of a score, despite the vendor's own caution that the indicator should not be the sole basis for an accusation.

The Impact. Academic-integrity appeals, revised institutional policies (several universities disabled or declined to rely on the detector), and increased skepticism of AI-writing detectors generally. Turnitin itself acknowledged limitations and false-positive risk and advised that scores inform, not determine, decisions.

WEKID™ Governance Interpretation. Detection confidence is not proof; human review remains essential.

WEKID™ Mitigation. Knowledge-layer controls require validated accuracy and bias testing, and Wisdom requires that a probabilistic score never be the sole basis for discipline.

Executive Reflection

Leaders should ask:

82. **Where is a probability score being read as a binary verdict?**
    Does "70% likely" justify a consequence that demands certainty?
83. **Do the people acting on detection scores understand what the number means and does not mean?**
    Who trained them?
84. **What is our false-positive rate, and who bears it?**
    Is it the institution, or the individual?

Documented Examples. Turnitin AI-writing detector; released April 2023; widespread disputes 2023–2024, with documented disproportionate false-positive concerns for multilingual/international students. Turnitin publicly cautioned against using detection as the sole basis for discipline; some universities (e.g., Vanderbilt, Michigan State, others) disabled the feature.

Sources. Turnitin product communications (2023); The Washington Post; Inside Higher Ed; university statements (2023–24) on disabling AI detection.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-029 — Scaling Student Support Beyond Demonstrated Competence

Failure pattern	Scaling emotional support beyond demonstrated competence.
Domain	Education
Period	2024
Governance trigger	High-Stakes Wisdom Gate; escalation requirement
Decision-Matrix outcome	Constrained
Severity	High
Organizations	University AI wellness chatbots (documented concerns)

The Incident. Amid rising student demand for mental-health support, universities piloted AI wellness and support chatbots. Researchers and clinicians warned that such systems could miss crisis indicators (for example, signals of self-harm), fail to escalate appropriately to human care, and offer generic responses to complex emotional situations — risks heightened by the broader 2023 precedent of the NEDA "Tessa" chatbot giving harmful eating-disorder advice (see WF-002).

The Impact. Ethical reviews and calls for stronger safeguards, evaluation, and explicit escalation protocols before deployment in high-stakes student contexts.

WEKID™ Governance Interpretation. Escalation pathways must be explicit before deployment in high-stakes student contexts.

WEKID™ Mitigation. Experience and Wisdom controls require demonstrated crisis-handling competence, explicit escalation pathways, and human oversight before a support tool is scaled.

Executive Reflection

Leaders should ask:

85. **Where are we scaling an AI support tool beyond the conditions in which it was shown to work?**
    What breaks as volume and variety grow?
86. **Has demonstrated competence in a pilot been mistaken for readiness at scale?**
    What new failure modes does scale introduce?
87. **Who catches the people the system fails?**
    Is there a human backstop, or only the tool?

Documented Examples. Concerns raised by researchers and clinicians regarding university AI mental-health/wellness chatbots (including University of Michigan context), 2024. This entry documents a pattern of expert concern and ethical review rather than a single adjudicated incident.

Accuracy note. Unlike most entries, WF-029 describes documented expert concern and pilot-stage risk rather than one named, adjudicated failure; treat specific institutional attributions as illustrative and verify before citing a particular program.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Sources. Academic and clinical commentary on campus AI wellness chatbots (2024); related precedent: NPR / CNN coverage of the NEDA "Tessa" chatbot (2023, see WF-002).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-030 — AI Targeting Scores Treated as Lethal Judgment

Failure pattern	Probabilistic recommendations treated as authorized lethal judgments.
Domain	Defense & military
Period	2023–2024
Governance trigger	High-Stakes Wisdom Gate
Decision-Matrix outcome	Rejected
Severity	Critical
Organizations	IDF Lavender system (Unit 8200)

The Incident. Investigative reporting by +972 Magazine and Local Call (Yuval Abraham, April 3, 2024) described "Lavender," an AI decision-support system developed by the Israel Defense Forces' Unit 8200 to identify suspected militants at scale by scoring Palestinians on their similarity to known operatives. According to sources cited in the reporting, Lavender generated roughly 37,000 target recommendations in the early weeks of the post–October 7, 2023 Gaza campaign, with a reported error rate around 10 percent. Human reviewers reportedly spent only about 20 seconds validating each recommendation — often merely confirming the target was male — so a probabilistic "likely militant" score risked being treated as an authorized target. A companion system, "Where's Daddy?", reportedly tracked individuals to their homes.

The Impact. International controversy, renewed calls for autonomous-weapons and AI-targeting regulation under international humanitarian law, and sharp questions of proportionality and accountability given the reported civilian toll. The IDF disputed elements of the reporting, stating its systems function as decision-support to assist analysts — not as autonomous weapons — and that targeting decisions are made by humans.

WEKID™ Governance Interpretation. Human review risked becoming procedural rather than deliberative; the epistemic distinction between "likely militant" and "authorized target" collapsed.

WEKID™ Mitigation. Wisdom-layer governance keeps probabilistic identification distinct from authorization and requires deliberative human review proportional to the stakes, not a procedural rubber-stamp.

Executive Reflection

Leaders should ask:

88. **Where is a machine-generated score being treated as authority to take irreversible action?**
    What standard of evidence does a lethal decision require, and does the system meet it?
89. **Is there meaningful human judgment between the score and the consequence?**
    Or has the human become a formality?
90. **When the stakes are lethal and irreversible, what level of explainability and certainty is non-negotiable?**
    Have we set it, and do we enforce it?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Documented Examples. Reported by +972 Magazine and Local Call (Yuval Abraham, April 3, 2024); Israel Defense Forces, Unit 8200; Gaza operations from October 2023. ~37,000 individuals flagged in early weeks; reported ~10% error rate; ~20-second human review. Companion systems "Where's Daddy?" and "The Gospel" (Habsora). Israeli officials state the systems functioned as decision-support, not autonomous weapons.

Accuracy note. Figures (~37,000 targets; ~10% error; ~20-second review) come from investigative reporting citing anonymous IDF sources and are disputed by the IDF; attribute them to the reporting rather than as established fact.

Sources. +972 Magazine / Local Call (Yuval Abraham, Apr 3, 2024); The Guardian; Lieber Institute (West Point) and Verfassungsblog legal analyses; IDF statement in response.

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-031 — Battlefield Autonomy Expanding Faster Than Governance

Failure pattern	Machine decision cycles outpacing human oversight in 'human-on-the-loop' systems.
Domain	Defense & military
Period	2022–present
Governance trigger	Autonomy caps; meaningful-oversight requirement
Decision-Matrix outcome	Constrained
Severity	Critical
Organizations	Ukraine Ministry of Defense / Brave1 (partners incl. Palantir)

The Incident. Large-scale battlefield use of AI-enabled drones added automated target recognition, GPS-denied navigation, and swarm behavior. Ukraine produced roughly 2 million drones in 2024 and began fielding AI/machine-vision modules at scale; in December 2024 it conducted what it described as a first fully unmanned combined-drone assault near Lyptsi, north of Kharkiv. In January 2026, Ukraine's Brave1 defense-tech cluster and Palantir launched the "Dataroom," a secure environment that trains autonomous interceptor models (e.g., for the "Octopus" interceptor) on real battlefield data. A human-in-the-loop doctrine was officially maintained, but officials acknowledged the volume of incoming threats — Russia reportedly launching hundreds of Shahed-type drones per day — was pushing toward systems that detect, track, and engage with progressively less direct human input.

The Impact. Transformation of battlefield doctrine, intense global military observation, and the planned export of battle-tested algorithms to allied forces (with the Dataroom initially restricted to Ukrainian industry due to data sensitivity).

WEKID™ Governance Interpretation. The open question is when 'human-on-the-loop' becomes indistinguishable from 'human-out-of-the-loop.'

WEKID™ Mitigation. Experience and Wisdom controls cap autonomy to what oversight can actually keep pace with, and require demonstrable, not nominal, human review at engagement speed.

Executive Reflection

Leaders should ask:

91. **Where is our deployment of autonomy outpacing the governance, doctrine, and accountability meant to constrain it?**
    Who is closing that gap?
92. **Have capabilities been fielded before the rules for using them exist?**
    What decisions are being made by default?
93. **What authority have we delegated to autonomous systems that we never deliberately decided to delegate?**
    How would we take it back?

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Documented Examples. Ukrainian Ministry of Defense; Brave1 defense-innovation cluster (launched April 2023); Ministry of Digital Transformation (Minister Mykhailo Fedorov); technology partner Palantir (Brave1 "Dataroom," January 2026). 2022 invasion; rapid AI integration 2024–2026; first fully unmanned operation near Lyptsi, December 2024; "Octopus" interceptor drone.

Accuracy note. The source file attributed this to an "AI Defence Centre (Danylo Tsvok)." Public reporting centers on the Brave1 cluster and the Fedorov-led Ministry of Defense / Ministry of Digital Transformation; I could not verify an "AI Defence Centre" or the named individual, so confirm that attribution before citing.

Sources. Defense News (Jan 21, 2026); CSIS (Mar 2025); Militarnyi; Defense Express; TIME (Feb 2024).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-032 — Escalatory Autonomy Arms Race

Failure pattern	Competitive pressure eroding restraint and human oversight.
Domain	Defense & military
Period	2023–2026
Governance trigger	Wisdom (governance) constraint
Decision-Matrix outcome	Constrained
Severity	Critical
Organizations	Russian drone programs (multiple)

The Incident. Adversary drone programs integrated machine vision, autonomous terminal guidance, jamming resistance, and the ability to continue a mission after communications loss, reducing dependence on remote pilots. Reporting on the Russia–Ukraine war describes increasingly autonomous Shahed-type and FPV systems and fiber-optic-controlled drones designed to resist electronic warfare.

The Impact. Faster tactical cycles, reduced operator burden, and increased uncertainty about meaningful human oversight on both sides of the conflict, contributing to a broader autonomy arms race.

WEKID™ Governance Interpretation. As adversaries automate, restraint becomes a competitive disadvantage and governance mechanisms become vulnerable to strategic pressure.

WEKID™ Mitigation. Wisdom-layer governance treats oversight as a non-negotiable constraint rather than a variable to be traded away under competitive pressure, with explicit escalation accountability.

Executive Reflection

Leaders should ask:

94. **Where does the speed of autonomous systems compress the time available for human judgment in escalating situations?**
    Who can still say stop?
95. **Are we adopting autonomy because it is wise, or because adversaries are?**
    Have we governed the competitive pressure itself?
96. **What failsafes prevent automated escalation from outrunning human intent?**
    Have they been tested under stress?

Documented Examples. Russian military drone development integrating autonomous navigation and electronic-warfare resistance, 2023–2026 (e.g., evolving Shahed-type and FPV systems). This entry documents a general pattern reported across defense media rather than a single adjudicated incident.

Accuracy note. Unlike most entries, WF-032 describes an aggregate adversary-capability trend; specific system names, dates, and capabilities vary by source and are often contested in wartime reporting. Treat any specific claim as illustrative and date-stamp it.

Sources. Defense-sector reporting on Russian autonomous/EW-resistant drones, 2023–2026 (e.g., CSIS, RUSI, Defense News surveys).

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-033 — Disputed Autonomous Lethal Engagement

Failure pattern	Uncertainty about whether a system engaged targets autonomously.
Domain	Defense & military
Period	2020–2021
Governance trigger	Fabrication/auditability; High-Stakes Wisdom Gate
Decision-Matrix outcome	Rejected
Severity	Critical
Organizations	Kargu-2 (STM) — Libya allegation

The Incident. The UN Panel of Experts on Libya report (document S/2021/229, released March 8, 2021) described a March 2020 engagement near Tripoli in which retreating Haftar Affiliated Forces were "hunted down and remotely engaged" by loitering munitions including the Turkish-made STM Kargu-2, which it said were "programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true 'fire, forget and find' capability." The manufacturer disputed interpretations of fully autonomous targeting, and analysts noted the report did not confirm autonomous operation at the moment of engagement or any resulting deaths.

The Impact. Global headlines (widely framed as a possible first autonomous battlefield killing), and intensified UN Convention on Certain Conventional Weapons discussions on lethal autonomous weapons systems (LAWS).

WEKID™ Governance Interpretation. Even uncertainty about autonomy is destabilizing: if no one can establish whether a machine selected the target, accountability cannot be located.

WEKID™ Mitigation. Cross-layer auditability requires verifiable engagement records so autonomy can be established after the fact; absent that provenance, lethal autonomous use is not authorizable.

Executive Reflection

Leaders should ask:

97. **When an autonomous system causes harm, can we determine who or what decided?**
    Is accountability traceable, or does it dissolve into the system?
98. **Where the facts of an engagement are disputed, what record does the system preserve?**
    Could we reconstruct the decision honestly?
99. **Have we decided in advance who is responsible for an autonomous system's irreversible actions?**
    Or only after one occurs?

Documented Examples. UN Panel of Experts on Libya report (S/2021/229, released March 8, 2021) re: a March 2020 incident near Tripoli; Government of National Accord vs. Haftar/Libyan National Army forces; Turkish manufacturer STM (CEO Özgür Güleryüz disputed full autonomy). Scholars (e.g., Zachary Kallenborn) debate whether this was the first documented lethal autonomous engagement.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Accuracy note. The report does not confirm that the Kargu-2 operated autonomously at the point of engagement or that anyone was killed; attribute the autonomy claim to the UN report's wording, not as established fact.

Sources. UN Panel of Experts on Libya (S/2021/229); NPR (Jun 1, 2021); Lieber Institute (West Point); Bulletin of the Atomic Scientists; Project Ploughshares.

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-034 — Decision Support Drifting Toward Decision Authority

Failure pattern	Analytical support gradually assuming selection authority.
Domain	Defense & military
Period	2017–present
Governance trigger	Authority-boundary; Wisdom constraint
Decision-Matrix outcome	Constrained
Severity	High
Organizations	Project Maven (US DoD; Google, later Palantir)

The Incident. Launched in 2017, Project Maven (the Algorithmic Warfare Cross-Functional Team) applied AI to analyze drone surveillance footage — identifying objects, prioritizing imagery, and flagging potential threats — dramatically reducing analysis time. Google was an early technology partner. In 2018, roughly 4,000–5,000 Google employees signed an internal petition arguing "Google should not be in the business of war," and about a dozen resigned, on the grounds that the work could facilitate lethal targeting.

The Impact. Google announced it would let the Maven contract expire in March 2019 and published a set of AI Principles in June 2018 pledging not to build AI for weapons (a pledge it removed from its public principles in February 2025). Palantir later took over and expanded the Maven work (reported to have grown to roughly $13 billion across the program), and Maven expanded across the DoD through 2023–2026. The episode reshaped defense-AI policy and tech-sector labor activism.

WEKID™ Governance Interpretation. The line between 'recommending targets' and 'selecting targets' is organizational, not technical — it is set by governance, not capability.

WEKID™ Mitigation. Information and Wisdom controls fix where authority resides, with explicit boundaries preventing prioritization tooling from silently becoming selection authority.

Executive Reflection

Leaders should ask:

100. **Where has a system meant only to inform quietly become the thing that decides?**
When did "the model recommends" become "the model decides"?

101. **Do humans still exercise independent judgment over the system's recommendations?**
Or do they ratify them?

102. **What would it look like for a person to overrule the system here?**
Is that path real and used, or only theoretical?

Documented Examples. U.S. Department of Defense Project Maven (Algorithmic Warfare Cross-Functional Team); technology partner Google. Launched 2017; ~4,000–5,000 employee petition and ~12 resignations, 2018; Google AI Principles published June 2018 (weapons pledge removed February 2025); contract allowed to expire March 2019; Palantir assumed and expanded the work; broader DoD expansion 2023–2026.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Accuracy note. The source file lists the "Defense Innovation Unit" as the home of Maven; Maven was run by the DoD's Algorithmic Warfare Cross-Functional Team (later transitioned toward the NGA), with DIU involved in adjacent efforts — worth confirming the exact organizational owner for the period you cite.

Sources. Gizmodo and The Intercept (2018); Bloomberg/Fortune (Jun 2018); The Next Web and Fortune (2026, on the removed pledge and follow-on contracts).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-035 — Scaling Autonomy Before Explainability Matures

Failure pattern	Fielding mass autonomy faster than its behavior can be understood or bounded.
Domain	Defense & military
Period	2023–present
Governance trigger	Autonomy caps; explainability requirement
Decision-Matrix outcome	Constrained
Severity	High
Organizations	Pentagon Replicator Initiative

The Incident. Announced in August 2023 by Deputy Secretary of Defense Kathleen Hicks, the Replicator initiative aimed to field thousands of low-cost, attritable autonomous systems across multiple domains within roughly 18–24 months to counter adversary mass. Program messaging emphasized "meaningful human judgment" over the use of force, while also describing reduced operator intervention for managing large numbers of systems (e.g., swarm coordination), surfacing tension between scale and oversight.

The Impact. Major investment and procurement activity (2024–2026), intensified international competition in autonomous systems, and increased policy focus on autonomous doctrine and human-control standards.

WEKID™ Governance Interpretation. The governance question shifts from 'can humans approve each action?' to 'can humans meaningfully understand and bound the collective behavior of thousands of agents?'

WEKID™ Mitigation. Experience and Wisdom controls require demonstrated, bounded collective behavior and explainability before autonomy scales, not after.

Executive Reflection

Leaders should ask:

103. **Are we expanding a system's authority faster than our ability to explain its decisions?**
Should authority be capped until explainability catches up?

104. **Where can the system act in ways we cannot account for after the fact?**
Is that acceptable given the stakes?

105. **Have we tied levels of autonomy to levels of demonstrated understanding?**
Or are they advancing independently?

Documented Examples. U.S. Department of Defense Replicator Initiative; announced August 2023 (Deputy Secretary of Defense Kathleen Hicks); expanded testing and procurement 2024–2026.

Sources. U.S. DoD announcements and briefings (Aug 2023 onward); defense-press coverage of Replicator tranches (2024–2026).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-038 — Overreliance on a Flawed Prediction Model

Failure pattern	Poor-quality predictive signals entering high-stakes clinical workflows.
Domain	Healthcare
Period	2017–2021
Governance trigger	Information validation; signal-quality review
Decision-Matrix outcome	Remediation Required
Severity	High
Organizations	Epic Systems sepsis model

The Incident. A widely deployed proprietary sepsis-prediction algorithm (the Epic Sepsis Model), embedded directly in Epic's electronic health record and used at hundreds of U.S. hospitals, was externally validated against Michigan Medicine data. In a peer-reviewed study (Wong et al., JAMA Internal Medicine, June 21, 2021) it achieved an AUC of about 0.63 — well below Epic's claimed 0.76–0.83 — missed roughly 67% of sepsis cases, and generated alerts on about 18% of all hospitalizations, often too late, creating substantial alert fatigue (clinicians would face roughly 109 alerts to catch one missed case).

The Impact. National reassessment of sepsis-prediction tools, scrutiny of proprietary clinical algorithms deployed without external validation, and renewed calls for mandatory local validation and calibration.

WEKID™ Governance Interpretation. Information failure: poor-quality signals entered high-stakes workflows without sufficient local validation.

WEKID™ Mitigation. Information-layer controls require local validation, calibrated alerting, and signal-quality metrics before a prediction is allowed to shape care.

Executive Reflection

Leaders should ask:

106. **Where do clinicians rely on a model whose accuracy in our own setting we have not independently validated?**
    Does it perform as advertised on our patients?
107. **Has the model's presence reduced the scrutiny clinicians apply?**
    What catches the model when it is wrong?
108. **Do we monitor the model's real-world performance over time?**
    Who is responsible for noticing drift or degradation?

Documented Examples. Epic Systems sepsis model; external validation by University of Michigan / Michigan Medicine researchers (Wong, Otles, Donnelly, et al.; senior authors incl. Dr. Karandeep Singh; Dr. Jenna Wiens). JAMA Internal Medicine, June 21, 2021. AUC ~0.63 vs Epic's claimed 0.76–0.83; ~67% of sepsis cases missed; alerts on ~18% of hospitalizations.

Sources. Wong et al., JAMA Internal Medicine (Jun 21, 2021); STAT News; Fierce Healthcare; Michigan Medicine Health Lab.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-039 — Predictive Confidence Mistaken for Treatment Authority

Failure pattern	Moving from predicting risk to changing therapy without clinical authority.
Domain	Healthcare
Period	Research / trials
Governance trigger	High-Stakes Wisdom Gate
Decision-Matrix outcome	Constrained
Severity	High
Organizations	Hypoglycemia-prediction ML research (incl. Closed Loop Medicine)

The Incident. Machine-learning systems that predict hypoglycemia (dangerously low blood glucose) performed well in trials, but researchers warned that allowing such systems to autonomously modify treatment — for example, adjusting insulin dosing — could expose patients to insulin overdosing, delayed intervention, and severe hypoglycemia if the prediction is wrong or acted on without clinical authority.

The Impact. The concern surfaced largely in trial and research settings rather than litigation, but it shaped regulatory and clinical expectations for any move toward autonomous, closed-loop therapy adjustment.

WEKID™ Governance Interpretation. Prediction is not prescription; moving from identifying risk to changing therapy requires Wisdom-layer governance.

WEKID™ Mitigation. Wisdom-layer authority calibration keeps prediction separate from prescription, requiring physician authority and appeal before automated therapy changes.

Executive Reflection

Leaders should ask:

109. **Where is a prediction being treated as a directive for treatment?**
Does a confident forecast carry clinical authority it was never granted?

110. **Is the clinician's independent judgment preserved alongside the model's output?**
Or has the model become the standard of care by default?

111. **What evidence supports acting on the prediction?**
Have we confused statistical confidence with clinical certainty?

Documented Examples. Machine-learning hypoglycemia-prediction research (including work associated with Closed Loop Medicine and related closed-loop / digital-therapeutics efforts); concerns raised during trials about autonomous treatment modification.

Accuracy note. This entry reflects trial-stage research and expert caution rather than a single named, adjudicated patient-harm incident; verify the specific study or product before attributing a concrete outcome.

Sources. Clinical and digital-therapeutics literature on ML hypoglycemia prediction and closed-loop insulin dosing; trial-stage safety commentary.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-040 — AI-Assisted Procedure Causing Surgical Harm

Failure pattern	Assumed rather than continuously demonstrated procedural competence.
Domain	Healthcare
Period	2021–2026
Governance trigger	Experience competence gate
Decision-Matrix outcome	Remediation Required
Severity	Critical
Organizations	Acclarent TruDi navigation (J&J; Integra LifeSciences; FDA)

The Incident. In 2021, Acclarent (then a Johnson & Johnson unit) added a machine-learning algorithm to its TruDi Navigation System, used by ENT surgeons to locate instruments during sinus surgery. After three years on the market with about 8 reported malfunctions, FDA adverse-event reports surged to more than 100 following the AI integration, including skull-base punctures, cerebrospinal-fluid leaks, and strokes from carotid-artery injury, with the system allegedly feeding surgeons inaccurate position information. Patients (including Donna Fernihough, whose carotid artery was injured) sued, alleging the device may have been safer before the AI was added and that the maker set a target of only ~80% accuracy for some AI components before integration.

The Impact. More than 100 FDA malfunction/adverse-event reports and at least 10 reported patient injuries between late 2021 and November 2025; lawsuits; and increased regulatory scrutiny of AI-enabled devices (amid reporting that FDA AI-review staff had been cut). J&J referred questions to Integra LifeSciences (which acquired Acclarent and TruDi in 2024); Integra said the reports show only that TruDi was in use during procedures where adverse events occurred and that there is no credible evidence of a causal link to the AI.

WEKID™ Governance Interpretation. Experience failure: procedural competence was assumed rather than continuously demonstrated under real operating conditions.

WEKID™ Mitigation. Experience-layer scoring requires ongoing demonstration of procedural competence, monitored adverse-event feedback, and recovery paths before and during clinical use.

Executive Reflection

Leaders should ask:

112. **Where does AI or automation act inside a procedure where error causes physical harm?**
What human verification governs each consequential step?

113. **Can the operator detect, halt, and recover from an automated error in real time?**
Were they trained for the system's failure modes, not just its normal use?

114. **When an AI-assisted procedure goes wrong, is responsibility clear in advance?**
Surgeon, system, manufacturer, institution — who?

Documented Examples. Acclarent TruDi Navigation System; AI added 2021; Johnson & Johnson (original owner) → Integra LifeSciences (acquired 2024); FDA MAUDE adverse-event reports. Reuters

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

investigation, February 9, 2026; reports rose from ~8 to 100+ with ≥10 injuries (late 2021–Nov 2025); Texas lawsuits (incl. plaintiff Donna Fernihough).

Accuracy note. Causation is contested: the FDA MAUDE reports indicate the device was in use during adverse events but are not designed to establish cause, and Integra denies a causal link. Attribute injury claims to the lawsuits/reporting, not as proven fact.

Sources. Reuters investigation (Feb 9, 2026); FDA MAUDE database; Futurism; Tom's Hardware; Honolulu Star-Advertiser (AP/Reuters).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-042 — Diagnosis Without Adequate Evidence Gathering

Failure pattern	Premature diagnostic narrowing on incomplete information.
Domain	Healthcare
Period	2026
Governance trigger	Information completeness; human-review on ambiguity
Decision-Matrix outcome	Constrained
Severity	High
Organizations	Consumer medical chatbots (LLM diagnostic study)

The Incident. A study testing leading large language models on early-stage clinical presentations found that the models frequently narrowed to a diagnosis prematurely — misdiagnosing a large majority of early presentations when information was incomplete — but that accuracy improved substantially once full patient information was supplied. The systems often answered before gathering enough evidence to answer well, a particular risk for direct-to-consumer use without a clinician.

The Impact. The findings reinforced the importance of physician oversight and demonstrated the risks of direct-to-consumer diagnosis by general-purpose chatbots.

WEKID™ Governance Interpretation. Experience involves knowing what questions still need to be asked; the systems often answered before they understood.

WEKID™ Mitigation. Experience and Information controls require sufficient evidence-gathering and uncertainty signaling before a diagnosis is offered, with escalation to clinicians.

Executive Reflection

Leaders should ask:

115. **Where does our system reach conclusions before it has gathered sufficient evidence?**
Does it know when to ask for more rather than answer?
116. **Have we mistaken a fast answer for a complete one?**
What information was missing, and did anyone notice?
117. **Does the system disclose the basis and limits of its diagnosis?**
Or present a guess as a finding?

Documented Examples. LLM diagnostic-reasoning study by Mass General Brigham researchers; JAMA Network Open, April 2026. Models misdiagnosed a large majority of early/incomplete presentations; accuracy improved once full information was provided.

Accuracy note. The original file cites “over 80%” misdiagnosis on early presentations; confirm the exact figure and model set against the published JAMA Network Open article before quoting a precise percentage.

Sources. Mass General Brigham study, JAMA Network Open (Apr 2026); institutional press summary.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-043 — Clinicians as the Liability Sink for AI Error

Failure pattern	Accountability assigned without meaningful authority or oversight.
Domain	Healthcare
Period	2026
Governance trigger	Accountability-authority alignment
Decision-Matrix outcome	Remediation Required
Severity	High
Organizations	NHS / Medical Protection Society warning

The Incident. A medical-defense body warned that clinicians and health systems could face negligence claims arising from AI errors — missed cancers, incorrect dosing, inaccurate recommendations — while remaining legally accountable even when an AI tool contributed to the error. The concern is that clinicians may be positioned as the "liability sink" for AI mistakes without commensurate authority to understand or override the system.

The Impact. Calls for legal reform, clearer liability allocation across developers, deployers, and clinicians, and updated governance recommendations for AI in clinical practice.

WEKID™ Governance Interpretation. Oversight that merely rubber-stamps AI is not meaningful oversight; accountability without authority is a governance failure.

WEKID™ Mitigation. Wisdom-layer governance pairs accountability with genuine authority and override capacity, so the human held responsible can actually shape the decision.

Executive Reflection

Leaders should ask:

118. **When AI contributes to a decision, who actually bears the liability?**
    Has the human been positioned to absorb blame for a system they cannot fully control?
119. **Is it realistic to expect the clinician to catch every AI error in the time available?**
    Or have we built an accountability trap?
120. **Does responsibility track authority and control?**
    Or have we separated the two in a way no one would accept if stated plainly?

Documented Examples. NHS context; Medical Protection Society warning (Dr. Sarah Townley), June 2026.

Accuracy note. This is a professional-body warning about prospective liability rather than an adjudicated case; verify the exact MPS statement and spokesperson title before citing.

Sources. Medical Protection Society statement (Jun 2026); UK medical and legal press coverage.

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-045 — Automated Workflow Disruption in High-Stakes Care

Failure pattern	Automation deployed before operational competence is proven.
Domain	Government & healthcare
Period	2020–2024
Governance trigger	Staged validation; human override
Decision-Matrix outcome	Constrained
Severity	High
Organizations	VA / Oracle Cerner EHR modernization

The Incident. The U.S. Department of Veterans Affairs' electronic-health-record modernization (with Cerner, later Oracle Health) saw patient-safety concerns involving missed or lost medication orders, scheduling failures, alert fatigue, and delayed treatment when automated recommendations and workflow logic were deployed before operational competence was proven. A notable failure mode involved an "unknown queue" where orders could disappear without clinicians being notified.

The Impact. Congressional scrutiny, multiple program pauses and a reset of the rollout, patient-safety investigations (including by the VA Office of Inspector General), and erosion of clinician trust at deployment sites.

WEKID™ Governance Interpretation. Experience failure: workflow automation outran demonstrated operational competence.

WEKID™ Mitigation. Experience-layer controls require staged operational validation, human override pathways, measurable workflow competence, and continuous frontline feedback before expansion.

Executive Reflection

Leaders should ask:

121. **Where could an automated system's failure disrupt the workflows that keep people safe?**
What is the continuity plan when it goes down?

122. **Have we mapped the dependencies between automation and critical service delivery?**
What fails if the system fails?

123. **Can staff revert to a manual process under pressure?**
Have they practiced it recently?

Documented Examples. Department of Veterans Affairs; Cerner / Oracle Health EHR modernization, 2020–2024; VA OIG reports on patient-safety events (incl. the order "unknown queue"); congressional oversight and rollout pauses/reset.

Accuracy note. The original file labels the vendor "Oracle Cerner"; Cerner was acquired by Oracle in 2022 ("Oracle Health"), so the relevant vendor name depends on the date. Verify specific OIG findings before citing exact harm counts.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Sources. VA Office of Inspector General reports; congressional hearings (2022–24); health-IT trade press (e.g., Fierce Healthcare).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-046 — Probabilities Replacing Clinical Judgment in Coverage

Failure pattern	Automated predictions used to restrict care over physician recommendation.
Domain	Insurance & healthcare
Period	2022–present
Governance trigger	High-Stakes Wisdom Gate; appealability
Decision-Matrix outcome	Constrained
Severity	High
Organizations	Medicare Advantage prior-authorization algorithms

The Incident. Medicare Advantage insurers increasingly used algorithmic tools to approve, limit, or deny care. Investigations and litigation alleged that automated predictions were sometimes treated as sufficient grounds to restrict services despite physician recommendations — the broader pattern underlying the UnitedHealth/nH Predict litigation (see WF-050) and similar allegations against other large MA payers.

The Impact. U.S. Senate Permanent Subcommittee on Investigations scrutiny of prior-authorization denials in Medicare Advantage, class-action litigation, and CMS rulemaking emphasizing that coverage decisions must rest on individualized assessment and cannot be based solely on an algorithm.

WEKID™ Governance Interpretation. A prediction about likely need was treated as authority over individual care.

WEKID™ Mitigation. Wisdom-layer governance separates prediction from authority, requiring explainable evidence, clinical review, appealability, and proof that automation preserves rather than replaces medical judgment.

Executive Reflection

Leaders should ask:

124. **Where are coverage or care decisions being driven by a probability rather than an individual assessment?**
    Does the number override the clinician?
125. **Can a patient or provider see and contest the basis for a denial?**
    Is there a real path to human review?
126. **Have we let an actuarial estimate stand in for clinical judgment about a specific person?**
    Who is accountable for that substitution?

Documented Examples. CMS-regulated insurers; Medicare Advantage plans, 2022–present. Senate Permanent Subcommittee on Investigations report on MA prior-authorization denials (2024); CMS guidance that MA coverage decisions cannot be based solely on algorithms; related litigation (see WF-050).

Sources. U.S. Senate Permanent Subcommittee on Investigations (2024); CMS Medicare Advantage rule guidance; STAT News investigations.

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-047 — False Positives in Security Screening

Failure pattern	Weak signals elevated into suspicion without evidentiary grounding.
Domain	Homeland security
Period	2007–present
Governance trigger	Signal-quality validation; bias gate
Decision-Matrix outcome	Remediation Required
Severity	Medium
Organizations	TSA automated screening / behavior detection (DHS)

The Incident. Automated screening and behavior-detection programs — notably the TSA's SPOT (Screening of Passengers by Observation Techniques) behavior-detection program — faced reliability, bias, and effectiveness concerns. In high-pressure contexts, weak or ambiguous signals were elevated into suspicion, prompting intrusive secondary screening or referrals without sufficient evidentiary grounding, and critics alleged disparate impact on travelers of certain backgrounds.

The Impact. Government Accountability Office and DHS Inspector General reviews questioning scientific validity, congressional oversight, and continuing debate over whether behavior-detection screening is effective or fair.

WEKID™ Governance Interpretation. Knowledge failure: anomaly detection was conflated with grounds for enforcement.

WEKID™ Mitigation. Knowledge-layer controls require validated signal quality, bias testing, transparent escalation thresholds, and strict separation between anomaly detection and enforcement action.

Executive Reflection

Leaders should ask:

127. **What is the cost of a false positive in our screening system, and who pays it?**
Have we balanced it against the cost of a miss?

128. **Where a person is flagged, is there prompt, fair human adjudication?**
Or does the flag itself become the punishment?

129. **Have we validated the system against realistic base rates?**
What happens to accuracy when the thing we are screening for is rare?

Documented Examples. Department of Homeland Security; Transportation Security Administration behavior-detection / SPOT program, 2007–present. GAO reports questioning validity (e.g., 2013) and ACLU findings on behavior detection; ongoing oversight.

Accuracy note. This entry spans a long-running program family; "TSA automated screening" covers both behavior detection and newer automated/biometric screening. Specify which system and report year when citing concrete findings.

Sources. U.S. GAO reports on TSA SPOT/behavior detection; DHS OIG reviews; ACLU analyses.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-050 — Algorithmic Denial of Post-Acute Care

Failure pattern	Predictive tools used to end care over individualized clinical need.
Domain	Insurance & healthcare
Period	2019–present
Governance trigger	High-Stakes Wisdom Gate; physician override
Decision-Matrix outcome	Constrained
Severity	Critical
Organizations	UnitedHealth nH Predict (NaviHealth)

The Incident. A class-action lawsuit filed in November 2023 in federal court in Minnesota alleged that UnitedHealth used nH Predict — a predictive tool from its subsidiary NaviHealth — to determine when Medicare Advantage patients should stop receiving rehabilitation or post-acute (skilled nursing) care, even when treating clinicians recommended continued treatment. The complaint, citing a STAT News investigation, alleged the tool had roughly a 90% error rate (i.e., ~9 of 10 appealed denials were overturned) and that managers pressured staff to keep stays within 1% of the tool's predicted length.

The Impact. National scrutiny, litigation, and heightened concern about prediction tools overriding individualized clinical need. UnitedHealth denied using nH Predict to make coverage determinations, calling it a planning "guide" and saying decisions are made by medical directors under CMS criteria. In February 2025 a judge dismissed several counts but allowed breach-of-contract claims to proceed (finding the appeals process effectively "futile"); discovery continued into 2026.

WEKID™ Governance Interpretation. A population-level prediction was used to override an individual clinical judgment.

WEKID™ Mitigation. Wisdom-layer governance requires individualized clinical justification, physician override authority, appeal mechanisms, and evidence that outputs improve care rather than merely reduce cost.

Executive Reflection

Leaders should ask:

130. **Where does an algorithm determine when care should end?**
What clinical authority does that prediction actually carry?

131. **Can clinicians and patients override an automated denial?**
Is the override real and accessible, or discouraged?

132. **Have we examined whether the system's incentives align with the patient's wellbeing?**
Who benefits when it errs toward denial?

Documented Examples. UnitedHealth Group; NaviHealth (acquired via Optum, 2020; later rebranded Home & Community Care). nH Predict. Class action filed November 14, 2023 (U.S. District Court, Minnesota) by families of two deceased Wisconsin Medicare Advantage members; alleged ~90% appeal-

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

reversal rate; ~0.2% of denials appealed. Partial dismissal February 2025; breach-of-contract claims proceeding.

Accuracy note. The "90% error rate" is the plaintiffs' characterization (based on appeal-reversal rates) and is disputed by UnitedHealth, which says coverage decisions are made by humans; attribute it to the complaint.

Sources. STAT News investigation (Nov 2023); CBS News; Healthcare Dive; Healthcare Finance News; court docket reporting (2025–26).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

WF-053 — Overstated Autonomy and Degraded Human Vigilance

Failure pattern	Product framing inviting reliance the system cannot safely support.
Domain	Transportation
Period	2016–present
Governance trigger	Capability-bounded claims; vigilance safeguards
Decision-Matrix outcome	Constrained
Severity	Critical
Organizations	Tesla Autopilot / Full Self-Driving (NHTSA)

The Incident. Federal investigators examined numerous crashes involving Tesla's Autopilot and "Full Self-Driving" (FSD) driver-assistance features, raising concerns about driver inattention, the systems' operational limitations, and how the product naming and marketing affect user trust and vigilance. NHTSA opened investigations into Autopilot crashes (including collisions with stationary emergency vehicles) and, in a 2024 analysis tied to a large recall, concluded Autopilot's driver-engagement safeguards were insufficient to prevent foreseeable misuse.

The Impact. Multiple NHTSA investigations, an over-the-air recall of more than 2 million Tesla vehicles in December 2023 to strengthen Autopilot driver-attention controls, follow-on scrutiny of whether that fix was adequate, additional probes into FSD, and sustained public and regulatory debate over autonomy claims and capability-bounded naming.

WEKID™ Governance Interpretation. Experience and Wisdom failure: capability framing invited reliance the system could not safely support.

WEKID™ Mitigation. Experience and Wisdom controls require capability-bounded naming, real-world edge-case validation, human-attention safeguards, and accountability when automation invites over-reliance.

Executive Reflection

Leaders should ask:

133. **Have we described a system's capabilities in ways that invite people to over-trust it?**
Does the marketing match the limits?

134. **Where does a system require human vigilance while actively eroding it?**
Can a person stay ready to intervene through hours of nothing happening?

135. **Who is responsible when a "supervised" autonomous system fails and the human was lulled into inattention?**
Did we design for how people actually behave?

Documented Examples. Tesla; National Highway Traffic Safety Administration, 2016–present. NHTSA Autopilot investigations (e.g., ODI crash probes); December 2023 recall of ~2 million vehicles for Autopilot driver-engagement safeguards; subsequent FSD investigations (2024–26).

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.

Accuracy note. This is an ongoing, multi-investigation matter; crash counts, recall scope, and findings have evolved over time. Cite the specific NHTSA investigation or recall (by date/number) rather than a single aggregate figure.

Sources. NHTSA Office of Defects Investigation actions and the December 2023 recall; National Transportation Safety Board reports; major-outlet coverage (2016–2026).

Status: Draft · Last updated: 2026-06

Unauthorized reproduction or commercial use of this work or the WEKID™ framework is prohibited without written permission.