



Ethics at the Edge: Decision Rights When AI Goes Off-Nominal

Artificial intelligence is rapidly becoming a critical agent, and sometimes an autonomous (agentic) actor inside the world's most critical systems: energy grids and water systems, transportation and logistics, healthcare and insurance, financial markets and enterprise operations. As the use of AI expands, and authorities provided AI systems grow, boards, executives and regulators face a parallel challenge: how should we govern / guard rail intelligent systems while empowering efficiency, monitoring insights and innovationⁱ? This is the problem space behind "Ethics at the Edge: Decision Rights When AI Goes Off-Nominal," a panel exploring the governance, legal, and operational dilemmas that arise when AI deviates from expected behavior, sometimes going 'off-nominal' (rogue).

Defining AI Opportunity and Risk

Many conferences and articles have addressed the need to build AI governance – policies, risk appetite, guard rails of permitted/prohibited roles – before or at least in parallel with expanding AI systems and rolesⁱⁱ. These actions will help define what qualifies as an AI error / anomaly (hallucination, bias, deviation) or AI emergency (critical event that requires action). AI *anomalies* happen daily - noise in a dataset, a model responding oddly to an edge condition – and detection, remediation and restructuring must be part of ongoing AI governance. But an AI *emergency* begins when autonomous actions create risks to safety, compliance, reliability, financial integrity or service quality. Boards must draw these lines explicitly, because the definition determines escalation authority and the latitude given to human operatorsⁱⁱⁱ.

Decision Rights in the Millisecond Domain

Traditional governance assumes human deliberation. AI upends this by acting faster than human reaction time. When decisions occur at the speed of computation, organizations need a well-architected authority ladder. The most proximate operational leader – in electric utilities, the cybersecurity on-call, control-room operator, or grid engineer - must have the authority to halt operations immediately. Executives must receive rapid notification, not pre-approval. Boards must own policy-level oversight, ensuring that authority is neither so centralized that people hesitate nor so distributed that no one acts^{iv}.

Kill-Switch and Pause Moment Governance

Kill switches have re-emerged as essential governance tools. Not the cinematic “big red button,” but multi-layered fail-safes: pause authority (when errors are detected, decisions are made and automated activities are de-escalated^v), hardware interlocks, software watchdogs, operator overrides, and system-wide shutdowns requiring dual authorization. These mechanisms must be governed with rigor: auditable logs, traceability, criteria for activation, and clear after-action review processes. When AI plays a safety-critical role, a kill-switch strategy is foundational^{vi}.

Preparing for the Off-Nominal Moment and Model Failure

Scenario design, tabletop exercises, red-team testing, and simulations are required elements of a robust governance playbook, as they expose ambiguous decision rights, undocumented technical dependencies and vendor blind spots. Not only do these exercises build muscle memory across functions - engineering, risk, legal, operations, communications^{vii} - but the likelihood of failures increases with cyberattacks. Security firm Zscaler recently reported that, when “full adversarial scans are run, critical vulnerabilities surface within minutes” (90% of AI systems within 90 minutes). Zscaler recommends that “organizations focus on visibility, real-time defense and consistent governance controls,” as well as scenarios and testing.^{viii}

The Ethical and Legal Dimension – including Vendor Accountability and Auditability

AI emergencies have ethical consequences: AI may optimize efficiency in ways that reduce fairness, service equity, or environmental constraints. The board must pre-encode ethical boundaries into operational systems, ensuring that when optimization pushes into ethically impermissible territory, the system automatically pauses or escalates to human review^{ix}. AI also has significant (and growing) legal consequences: disclosure obligations, contract exposures, liability allocation with vendors, and regulatory oversight. A key board oversight role is ensuring management monitors, complies with, and reports on shifting regulatory requirements, not just in the home country, but in jurisdictions where customers, vendors and partners operate^x. Complementing this regulatory challenge, however, are some fit-for-purpose AI tools to monitor laws and regulations, conflicts of interests, sanctions and trade compliance, revenue recognition and partner risks, fraud and waste, employee travel and entertainment, 3rd party risks and improper payments, anti-bribery and corruption^{xi}. Boards and executives must be alert to AI-driven components embedded in vendor products, which complicates forensic analysis, responsibility, and control. Boards must require vendor transparency, access to logs, incident participation, and compliance with internal governance standards. Model explainability and auditability are conditions of deployment^{xii}.

Post-Incident Learning and Governance Refresh

The most valuable governance happens after an incident. A robust after-action process should include unified

timelines, cross-functional debriefs, vendor forensics, and measurable remediation plans. Boards must review these events not as failures but as essential inputs to governance maturity^{xiii}.

The Path Ahead

As AI becomes an operational co-pilot for mission-critical infrastructure, boards must evolve their oversight. Governance must extend from traditional risk frameworks into real-time decision architectures. Ethical decision rights, kill-switch strategies, auditability, and resilience are the pillars of next-generation AI governance. AI's failures will not always be predictable—but our preparedness can be.

ⁱ <https://www.covasant.com/hubfs/resources/white-papers/The-Key-Role-of-Agentic-AI-in-Fraud-Detection-and-Compliance.pdf>

ⁱⁱ <https://www.fortalicesolutions.com/post/innovation-ai-compliance-risk-assessments>

ⁱⁱⁱ <https://www.obsidiansecurity.com/blog/what-is-ai-governance>, <https://elevateconsult.com/insights/ai-governance-program-documentation-kit-for-legal-teams/>, <https://www.nist.gov/itl/ai-risk-management-framework>, <https://www.mineos.ai/articles/ai-governance-framework>, <https://www.modelop.com/ai-governance/ai-governance-roles>, <https://www.alation.com/blog/ai-governance-best-practices-framework-data-leaders/> and <https://www.ibm.com/think/topics/ai-governance>

^{iv} <https://www.coalitionforsecureai.org/defending-ai-systems-a-new-framework-for-incident-response-in-the-age-of-intelligent-technology/>, <https://www.snowflake.com/en/fundamentals/ai-governance-threat-intelligence-and-ml-anomaly-detection>

^v <https://www.humanisticpowerllc.com/hpg-ecosystem>

^{vi} <https://mindgard.ai/blog/ai-risk-management-framework>

^{vii} <https://www.cyber.gov.au/business-government/secure-design/operational-technology-environments/principles-for-the-secure-integration-of-artificial-intelligence-in-operational-technology>, <https://www.coalitionforsecureai.org/defending-ai-systems-a-new-framework-for-incident-response-in-the-age-of-intelligent-technology/>

^{viii} [https://www.linkedin.com/shareArticle?mini=true&url=https%3A//www.cybersecuritydive.com/news/ai-vulnerabilities-governance-zscaler/810718/&title=AI tools break quickly%2C underscoring need for governance](https://www.linkedin.com/shareArticle?mini=true&url=https%3A//www.cybersecuritydive.com/news/ai-vulnerabilities-governance-zscaler/810718/&title=AI%20tools%20break%20quickly%2C%20underscoring%20need%20for%20governance)

^{ix} <https://www.ai21.com/knowledge/ai-governance-frameworks/>, https://www.splunk.com/en_us/blog/learn/ai-governance.htm and <https://www.joneswalker.com/en/insights/blogs/ai-law-blog/ai-governance-series-part-3-building-governance-that-actually-works.html>

^x <https://www.morganlewis.com/pubs/2025/12/the-new-rules-of-ai-a-global-legal-overview>

^{xi} www.konaai.com tools.

^{xii} <https://www.nist.gov/itl/ai-risk-management-framework>, <https://oecd.ai/en/incidents>, <https://www.cisa.gov/topics/information-communications-technology-supply-chain-security> and <https://www.iso.org/standard/42001>

^{xiii} <https://witness.ai/blog/ai-compliance-framework>, <https://artificialintelligenceact.eu/>, <https://www.ntia.gov/page/software-bill-materials>