Beyond the Black Box: Integrating Legal Competencies and Algorithmic Liability into the Engineering Curriculum

Zsolt Czékmann[1]* and Gergely Cseh-Zelina[2]

[1] Department of Public Administrative Law, Faculty of Law, University of Miskolc, Miskolc, Hungary zsolt.czekmann@uni-miskolc.hu
[2] Department of Public Administrative Law, Faculty of Law, University of Miskolc, Miskolc, Hungary gergely.cseh-zelina@uni-miskolc.hu

Abstract. The rapid proliferation of autonomous artificial intelligence systems has precipitated a profound liability crisis that traditional engineering education is ill-equipped to address. Historically insulated by the "provided as-is" doctrine, software engineers now face an evolving geopolitical regulatory matrix in which algorithmic opacity, emergent bias, and unprompted decision-making can yield tangible legal and physical harms. This comprehensive report details the collapse of legacy liability shields and the emergence of strict, divergent regulatory empires across the European Union, the United States, and China. By synthesising recent jurisprudence - specifically, the algorithmic hallucinations in the Air Canada litigation and the fatal system safety failures in the Uber autonomous vehicle collision - this analysis identifies the precise intersections at which technical design choices become actionable legal negligence. To navigate this engineering-legal singularity, this report proposes a systemic restructuring of higher education curricula. It advocates cultivating "T-shaped" engineers by integrating Value-Based Engineering standards, implementing "Legal Autopsies" as a core pedagogical framework, establishing interdisciplinary legal-engineering clinics, and deploying blockchain-secured micro-credentials. Ultimately, this report demonstrates that legal literacy can no longer be relegated to the periphery of engineering education; it must be embedded as a fundamental boundary condition of algorithmic design.

Keywords: Algorithmic Liability, Engineering Education, EU AI Act, Legal Autopsy, Value-Based Engineering, Artificial Intelligence Regulation, Micro-credentials.

1 The Engineering-Legal Singularity

1.1 The Collapse of the "Provided As-Is" Doctrine

For decades, the software engineering profession operated under a unique liability shield distinct from the rigorous accountability frameworks governing civil, structural,

© The Author(s) 2026
S. Árvai-Homolya et al. (eds.), *Proceedings of the 2nd Innovation in Engineering Education Conference (IEE 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 58,
https://doi.org/10.2991/978-94-6239-823-8_7

or mechanical engineering [1]. While a civil engineer could historically be held criminally and civilly liable for a bridge collapse resulting from structural negligence, software developers were largely insulated by End User License Agreements and the pervasive "provided as-is" doctrine. This legacy legal framework treated software errors as inevitable, probabilistic glitches rather than actionable negligence, provided there was no demonstrable malicious intent. Consequently, engineering curricula prioritised optimisation, efficiency, and computational elegance, treating legal compliance as a secondary, post-development administrative task rather than a core design constraint [2].

This era of overarching immunity has effectively ended with the widespread deployment of autonomous artificial intelligence agents and the integration of Large Language Models (LLMs) into high-stakes, consumer-facing roles. The deployment of these sophisticated systems has pierced the corporate veil that previously separated the individual engineer from the tangible physical, financial, and societal harms caused by their creations [3]. Unlike deterministic software, where fixed inputs yield strictly predictable outputs, modern artificial intelligence systems exhibit emergent behaviours -such as unprompted decision-making, latent bias, and algorithmic hallucinations -that directly intersect with fundamental individual rights and physical safety.

The judiciary, as evidenced by an accelerating volume of recent jurisprudence, is increasingly rejecting the "black box" defense [4]. The argument that an engineer or a deploying corporation cannot be held liable because a machine's internal decision-making pathway is mathematically opaque is no longer legally tenable. Instead, the legal system is aggressively reframing algorithmic opacity itself as a fundamental design defect. The engineer of the current decade must operate within a new paradigm in which legal code and software code are essentially isomorphic [5]. For example, a failure to implement a deterministic guardrail in a Retrieval-Augmented Generation (RAG) system is no longer merely a bug; it is increasingly recognised as a potential tort of negligent misrepresentation. Similarly, the failure to curate a training dataset free of historical bias is not merely poor data science; it constitutes a statutory violation of fundamental human rights under emerging regulatory regimes such as the European Union Artificial Intelligence Act (EU AI Act). Consequently, the engineering curriculum must rapidly evolve to treat legal compliance not as an ex-post-facto safety check but as a primary boundary condition of systemic design, carrying the same absolute weight as thermodynamic limits or material stress tolerances. Furthermore, shifting patterns of legal liability in both the private and public sectors necessitate that engineering students grasp the profound implications of deploying AI in critical domains, as traditional theories of fault-based liability are fundamentally challenged by distributed algorithmic causation.

The engineer of 2026 must operate within a new paradigm in which legal code and software code are isomorphic [6]. A failure to implement a "guardrail" in a Retrieval-Augmented Generation (RAG) system is no longer just a bug; it is a potential tort of negligent misrepresentation. A failure to curate a training dataset free of bias is not merely poor data science; it is a statutory violation of fundamental rights under regimes

such as the EU AI Act. Consequently, the curriculum must evolve to treat legal compliance not as an ex post facto check but as a primary boundary condition for design, equivalent to thermodynamic limits or material stress tolerances [7], [8].

1.2 The Geopolitical Regulatory Matrix

The global engineering landscape has fractured into distinct regulatory empires, each imposing fundamentally different, and frequently contradictory, constraints on artificial intelligence system design. An engineering graduate entering the contemporary workforce must be prepared to navigate these divergent realities, recognizing that the most computationally optimal engineering solution in one jurisdiction may be strictly illegal in another [9]. The lack of deep global harmonization means that multinational deployments require parallel compliance architectures or the adoption of the strictest framework as a de facto global ceiling.

The European Union: The Architecture of Pre-Compliance (Ex-Ante).

The European Union Artificial Intelligence Act represents the world's most structured, "engineering-first" regulatory framework, operating on a precautionary principle that requires engineers to categorise systems by objective risk before any development begins. This framework effectively creates a "negative design space," delineating specific areas where engineers are legally forbidden from venturing, regardless of technical feasibility or market demand. The regulation attempts to embody "smart regulation" by utilizing proportional ex-ante controls, targeting the highest risks with stringent measures while maintaining flexibility for minimal-risk applications. Crucially, the legislation imposes an "As Low As Reasonably Practicable" (ALARP) standard for residual risk, mandating that providers eliminate risks as far as technically feasible before relying on warnings or secondary controls [10].

For systems classified under unacceptable risk, the Act explicitly bans specific engineering functionalities. Systems employing subliminal techniques to distort human behavior, social scoring algorithms, or biometric categorization in public spaces are outright illegal. This strict prohibition requires the engineer to deeply understand that certain optimization goals, such as maximizing user engagement via subconscious psychological nudges, are legally out of bounds.

For high-risk architectures - such as those deployed in medical devices, critical infrastructure, or employment sorting - the Act imposes heavy engineering burdens that permanently transform standard software development practices. Data governance requirements dictate that training data must be relevant, representative, and strictly free of errors, a provision that elevates data cleaning from a heuristic task into a rigid compliance mandate. Engineers must employ advanced statistical methods to empirically demonstrate non-bias and representativeness, thereby generating a continuous documentation trail that explicitly links the foundational dataset to the final model's behavioural outputs. Furthermore, traceability and logging requirements essentially outlaw the "black box" in high-risk contexts. Architectures must support automatic, immutable recording of events to allow for rigorous post-incident forensic audits, a requirement that dictates specific architectural choices favoring interpretable models or highly ro-

bust logging layers over opaque, monolithic neural networks [11]. Transparency requirements necessitate that providers understand precisely "who must explain what" across the entire supply chain, establishing clear delineations of responsibility between foundational model developers and downstream deployers [12]. Crucially, the interface design must include mandatory human oversight, explicitly allowing a human operator to override the artificial intelligence—a requirement that extends beyond user interface design to become a legally mandated fail-safe.

The United States: Federal Preemption and Tort Liability (Ex-Post).

The regulatory landscape in the United States underwent a seismic shift with the issuance of Executive Order 14365 in December 2025, titled "Ensuring a National Policy Framework for Artificial Intelligence" [13]. This executive action reflects a strategic geopolitical pivot toward deregulation and federal preemption, explicitly designed to maintain American dominance in artificial intelligence innovation by removing burdensome localized constraints [14]. Before this order, engineers faced a highly complex patchwork of state-level regulations, including strict algorithmic bias laws in Colorado and comprehensive safety bills in California. Executive Order 14365 established an AI Litigation Task Force mandated to challenge state laws deemed inconsistent with federal policy, seeking to preempt these localized statutes in favor of a unified, less prescriptive federal standard. While Congress previously defeated a full ten-year moratorium on state AI laws, the executive branch's aggressive push toward preemption fundamentally altered the compliance landscape [15].

For the engineering practitioner, this environment creates a volatile paradox. While prescriptive, ex-ante safety rules have been heavily curtailed, this deregulation has drastically increased the industry's reliance on the ex-post tort system. Despite federal deregulation, the individual engineer and their deploying organization remain strictly subject to the "Duty of Care" under traditional tort law. If a deployed system causes physical or financial harm, the developers face severe litigation for negligence. In the absence of strict statutory guardrails, the judiciary looks to established "industry standards" to determine if the engineering team acted with reasonable professional care. Within this context, the National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework (RMF) becomes the engineer's primary legal shield. Rigorous compliance with its core functions—Govern, Map, Measure, and Manage—serves as critical evidence of reasonable care. Consequently, adversarial testing, or "Red Teaming," transitions from being merely a security best practice to an absolute legal necessity, where the failure to aggressively probe a model for vulnerabilities can directly constitute negligence in a liability lawsuit.

China: The Semantic Firewall (Content Control).

China’s regulatory approach imposes a distinct "truth and accuracy" constraint on algorithmic outputs that diverges significantly from Western regulatory models. Driven by a philosophy of social stability and state control, algorithms are legally mandated to uphold core socialist values and explicitly avoid generating false or disruptive information. This imposes a profound semantic constraint on the fundamental engineering of Large Language Models. Engineers cannot rely on simple heuristic keyword matching; instead, they must build highly sophisticated content filtering layers capable of

understanding deep semantic context to align the system with a specific, state-sanctioned political reality.

Furthermore, the implementation of an Algorithm Registry requires developers to fully disclose the basic logic, training data sources, and intervention mechanisms of their models to the Cyberspace Administration of China. This framework effectively mandates Explainable Artificial Intelligence (XAI) as a hard requirement for market entry. Engineers are stripped of the ability to hide behind the mathematical complexity of deep neural networks; they must be able to clearly explain the precise logic of an algorithmic decision to regulatory bodies, reflecting a highly centralised, state-driven approach to algorithmic governance. This requires embedding technical transparency measures directly into continuous integration pipelines in software development, transforming compliance from a legal abstraction into a strict engineering parameter.

1.3 Comparative Analysis for Curriculum Design

The divergence in engineering requirements among these three regulatory blocs necessitates a fundamental paradigm shift in the way global engineering curricula are structured [16]. Curriculum designers must synthesize these varying constraints to prepare students for international deployment realities, ensuring that graduates possess the nuanced legal literacy required to map technical parameters to distinct geopolitical requirements.

Table 1. Comparative Analysis of Global AI Regulatory Frameworks (EU, US, and China)

Feature	European Union (AI Act)	United States (Post-EO 14365)	China (Generative AI Measures)
Regulatory Philosophy	Precautionary / Safety-by-Design (Ex-Ante)	Innovation-First / Liability-Driven (Ex-Post)	Social Stability / Content Control
Primary Engineering Constraint	Risk Classification (The Risk Pyramid)	Duty of Care / Negligence Mitigation	Truthfulness & Value Alignment
Data Requirements	Representative, error-free datasets (Art. 10)	Focus on IP & Privacy (Litigation risk)	Lawful sources; Truthful content
Black Box Status	Transparency mandated for High-Risk systems	Permitted, but heavily increases the negligence risk	Must fully explain logic to state regulators
Human Role	Mandatory Human Oversight (HCI)	Human-in-the-loop as a liability shield	Provider is strictly responsible for the generated content
Key Compliance Artifact	Technical Documentation / CE Marking	NIST AI RMF Profile / Red Team Report	Algorithm Filing / Security Assessment

2 Anatomy of Algorithmic Failure – The "Legal Autopsy"

To effectively teach highly abstract legal and ethical concepts to technically minded students, engineering education requires concrete, dissectible examples of systemic failure. The "Legal Autopsy" method involves meticulously deconstructing a published legal judgment not merely to understand the underlying jurisprudence, but to isolate the root engineering cause. This process traces the path from a specific design decision, architectural configuration, or line of code directly to the precipitation of legal liability. By analyzing failures through this lens, students begin to perceive the direct causal relationship between algorithmic opacity, distributed causation, and real-world harm.

2.1 Case Study A: Moffatt v. Air Canada – The Hallucinating Agent

The intersection of generative algorithms and corporate liability was sharply defined by the landmark 2024 ruling by Canada's Civil Resolution Tribunal in Moffatt v. Air Canada. The plaintiff queried the airline's customer service chatbot regarding bereavement fares, and the system, operating on a Retrieval-Augmented Generation architecture, erroneously instructed the passenger that they could apply for a retroactive refund within a ninety-day window. Relying on this generated advice, the passenger purchased a full-price ticket, only to have the airline deny the subsequent refund claim, citing its static, official tariff policy that explicitly forbids retroactive bereavement discounts.

In its legal defense, the corporation attempted a novel argument, suggesting that the chatbot was a "separate legal entity" responsible for its own actions. The tribunal categorically rejected this defense as remarkable and legally baseless, ruling that an artificial intelligence agent is merely a component of the corporate digital infrastructure, and the deploying entity is strictly liable for negligent misrepresentation regardless of whether the false information is generated by a human employee or a probabilistic algorithm. The implications of this ruling echo deeply into corporate accountability, directly challenging classical fault-based liability models by confirming that corporations cannot use algorithmic autonomy to shield themselves from derivative liability [17].

Subjecting this case to an engineering autopsy reveals a classic, preventable breakdown in system architecture. The technical failure occurred because the retrieval mechanism successfully accessed relevant information but the generative model failed to prioritize the absolute ground truth of the static tariff document over its inherent probabilistic text generation capabilities. The fundamental design flaw was the absence of rigid, deterministic guardrails. A robustly engineered system deployed in a domain handling critical financial policies must include deterministic overrides that recognize sensitive queries—such as those regarding refunds or binding policies—and automatically revert to strict information extraction or refer the user to verified static pages, entirely bypassing the generative model. This case effectively dismantles the cognitive error common among engineers of anthropomorphizing algorithms, establishing firmly that the system possesses no agency and that liability rests entirely on the design and deployment choices made by the engineering team.

2.2 Case Study B: State of Arizona v. Uber – The Disabled Fail-Safe

The catastrophic consequences of prioritizing algorithmic tuning over fundamental system safety were exposed in the 2018 incident where an autonomous Uber test vehicle struck and killed pedestrian Elaine Herzberg in Tempe, Arizona. A detailed investigation by the National Transportation Safety Board revealed that the vehicle's perception systems, using both LIDAR and radar, detected the pedestrian approximately 6 seconds before impact. However, the vehicle systematically failed to initiate any emergency braking protocols.

The engineering autopsy of this fatality isolates the root cause not as a sensor failure or an unforeseeable edge case, but as a deliberate, conscious configuration choice made by the development team. In an effort to optimize ride smoothness and prevent "phantom braking"—a phenomenon where the vehicle brakes erratically for false positives such as steam or floating debris—engineers explicitly disabled the factory-equipped emergency braking system while the vehicle operated under computer control. The entire safety architecture was subsequently engineered to rely exclusively on a human safety driver to intervene in emergency scenarios.

This design completely ignored the well-documented psychological phenomenon of automation bias. When an automated system performs reliably for extended periods, human operators physiologically and psychologically lose situational awareness, shifting from active monitors to passive observers. In this incident, the safety driver was distracted by a mobile device, rendering the fundamental engineering assumption—that a human can instantaneously regain full situational awareness and execute evasive maneuvers within seconds—fatally flawed. The legal and ethical negligence lies in the conscious decision to disable a critical hardware safety feature to improve a software user-experience metric, violating the foundational engineering duty of care and demonstrating a profound failure in system safety engineering. This event highlights the complexities of distributed causation, where liability is scattered among software developers, corporate executives, and end-user operators, rendering traditional legal frameworks inadequate and exposing severe gaps in engineering education [18].

3 Pedagogical Paradigms for the AI Era

The unprecedented velocity of algorithmic technology development and the corresponding evolution of the regulatory environment necessitate a complete dismantling and restructuring of traditional engineering education. The legacy "siloed" academic model, wherein engineering and legal faculties operate in complete isolation, is functionally obsolete in an era where technical outputs are subject to immediate legal scrutiny. Emerging research indicates that higher education institutions must implement comprehensive regulatory frameworks internally to manage AI usage effectively, bridging the gap between technological capabilities and pedagogical integrity.

3.1 The "T-Shaped" Engineer: A New Competency Profile

The traditional "I-shaped" engineer—characterized by deep, specialized technical proficiency but lacking interdisciplinary breadth—is severely disadvantaged in the modern era of algorithmic regulation. To prevent the catastrophic failures detailed in the legal autopsies, academic institutions must pivot toward cultivating "T-shaped" engineers. The vertical bar of this competency profile represents an unwavering commitment to deep technical mastery; the engineer must maintain absolute expertise in neural network architectures, optimization algorithms, control theory, and advanced data science, as there is no substitute for technical rigor.

The critical evolution lies in the horizontal bar, which represents broad, cross-disciplinary fluency [19]. In the contemporary landscape, this breadth must encompass comprehensive legal literacy, specifically spanning data privacy frameworks like the General Data Protection Regulation, intellectual property law, and the nuances of algorithmic liability. The pedagogical objective is not to transform technical students into legal scholars, but to equip them with the analytical capacity to immediately recognize when a specific technical decision or architectural trade-off generates a material legal risk [20]. A legally literate engineer understands that "data minimization" is not merely an elegant storage optimization strategy, but a strict statutory requirement. As demonstrated by shifting global paradigms, such as the United Arab Emirates mandating comprehensive AI literacy across educational domains to combat algorithmic bias, cultivating this broad competency is essential for mitigating systemic risks.

Furthermore, this competency profile must actively integrate standardized ethical frameworks, most notably the IEEE 7000-2021 Standard Model Process for Addressing Ethical Concerns During System Design [21]. Value-Based Engineering (VBE) frameworks teach engineers how to systematically identify the values of all affected stakeholders, translating highly abstract ethical concepts such as "fairness" and "privacy" into explicitly measurable, testable system requirements [22] [23].

Table 2. IEEE 7000-2021 Standard Modell Process Phases and Engineering Purposes.

IEEE 7000-2021 Process Phase	Engineering Purpose
Value Elicitation	Identify ethical values relevant to the system and its stakeholders.
Value Prioritization	Resolve conflicts between competing values through structured stakeholder dialogue.
Ethical Risk Assessment	Identify and evaluate risks to values throughout the concept and design phases.
Values Traceability	Track ethical values through operational concepts, requirements, and design decisions.

By embedding these standards into the core curriculum, the T-shaped engineer learns to validate that implemented systems actively uphold prioritized values through continuous measurement, fundamentally bridging the gap between theoretical ethics and concrete software engineering.

3.2 Clinical Legal-Engineering Education (CLEE)

To operationalize this interdisciplinary competency, engineering faculties must adopt the highly successful clinical models utilized in medical and law schools by establishing Interdisciplinary Artificial Intelligence Clinics. These experiential learning environments structure collaboration between senior engineering students and law students, placing them in mixed teams to service real-world clients, such as university research groups, local startups, or non-profit organizations developing algorithmic tools.

Operating under Kolb's cycle of experiential learning—which emphasizes active experimentation followed by guided reflection—these clinics task students with executing comprehensive Algorithmic Impact Assessments and regulatory compliance audits. Within this collaborative framework, the law students identify the specific regulatory boundaries and legal obligations, determining, for example, if a client's diagnostic tool classifies as a high-risk system under the EU AI Act requiring strict human oversight. Simultaneously, the engineering students translate these legal mandates into precise technical specifications, designing confidence-interval dashboards and deterministic manual override switches to satisfy the legal requirement.

The successful implementation of these clinics requires adherence to a multilevel framework for AI integration, encompassing technological infrastructure, institutional support, and cultural adaptation. Utilizing models derived from the Technology Acceptance Model (TAM) and Social Cognitive Theory (SCT), educators can effectively foster an environment where students do not simply memorize legal statutes, but actively apply them through AI-inclusive assessment design and project-based learning. The resultant output is a joint, interdisciplinary report that simulates the cross-functional product development environments students will encounter in modern enterprise settings, effectively bridging the communication gap between legal counsel and development teams.

3.3 Micro-credentials and Lifelong Learning

The extraordinary velocity of artificial intelligence regulation renders a static, four-year degree program structurally insufficient. Knowledge regarding specific compliance frameworks acquired during a student's freshman year may be entirely legally obsolete prior to their graduation. To mitigate this epistemological decay, higher education institutions must develop and deploy architectures for stackable micro-credentials.

These micro-credentials consist of highly focused, intensive certification modules requiring twenty to forty hours of study, targeting highly specific, rapidly evolving intersections of law and technology. Topics must be aggressively relevant, covering subjects such as Algorithmic Bias Auditing, Generative Artificial Intelligence Copyright Compliance, and practical implementations of the NIST Risk Management Framework.

A systematic review utilizing the PRISMA methodology underscores that micro-credentials play a vital role in upskilling and reskilling the workforce, filling the immediate competency gaps left by rigid traditional academic structures. These modules are crucial not only for current undergraduate students but also serve as vital mechanisms for mid-career engineering professionals who graduated prior to the generative AI boom and now face unprecedented legal exposure.

To ensure the integrity, portability, and international recognition of these specialized competencies, universities should leverage blockchain technology for secure credentialing. By leveraging decentralised verification and self-sovereign identity models, academic institutions can issue tamper-proof digital certificates, often using protocols such as ERC-721 non-fungible tokens or automated smart contracts. This blockchain integration explicitly prevents the falsification of credentials—a systemic issue in remote learning—and allows employers to cryptographically verify that an engineer possesses the exact, up-to-date legal literacy required to develop compliant systems in high-risk environments. By anchoring these credentials on an immutable ledger, universities standardize trust in an increasingly decentralized educational ecosystem, fostering a culture of lifelong learning that tracks dynamically with regulatory shifts.

4 Case Studies in Detail

A deeper, highly technical analysis of algorithmic failures reveals exactly how traditional fault-based liability and strict liability concepts are currently failing to address the distributed causation and deep opacity inherent in modern machine learning architectures. Courts are actively struggling with the burden of proof, forcing an evolution in how liability is assigned and emphasizing the absolute necessity for meticulous engineering documentation and traceability [24].

4.1 Moffatt v. Air Canada: The Limits of Automated Agency

The case of Moffatt v. Air Canada provides a profound exploration of derivative liability and the legal boundaries of algorithmic autonomy. The plaintiff, seeking bereavement travel accommodations, engaged with the airline's automated interface, which unequivocally stated that a retroactive refund could be processed within ninety days of ticket issuance. Upon the airline's subsequent rejection of the refund request—which relied on the static, actual corporate tariff rules—the legal dispute centered on whether the corporation could sever its liability from the autonomous output of its digital agent.

During the tribunal proceedings, the corporate defense relied heavily on attempting to establish the chatbot as an independent, separate legal entity, arguing that the airline should not be held vicariously liable for the unprompted generation of false information by an algorithmic third-party tool. The Civil Resolution Tribunal dismantled this argument, holding that the interactive component of a website carries the same legal weight as a static text page. The corporation is strictly responsible for all data conveyed through its authorized channels.

From an engineering perspective, this case illustrates a fundamental failure in understanding the limitations of generative models. The phenomenon of "hallucination"—where a model generates syntactically plausible but factually incorrect information based on probabilistic token prediction rather than factual extraction—is a known architectural reality. The failure to implement deterministic boundaries around the model's capacity to communicate legally binding policies constitutes negligent system design. The engineering team failed to conduct adequate adversarial testing on policy-specific queries, demonstrating a lack of necessary legal literacy regarding the concept of negligent misrepresentation. Furthermore, this lack of oversight touches upon broader compliance issues, suggesting potential vulnerabilities concerning General Data Protection Regulation mandates if automated systems hallucinate or improperly disclose sensitive user data without rigid structural safeguards. The pedagogical value of this case is immense: it explicitly teaches students that algorithms are strictly tools of the deploying organization, possessing no independent legal agency, and that architectural design must aggressively prioritize verified extraction over generative probability in any domain involving financial or contractual commitments.

4.2 State of Arizona v. Uber: The Cost of Disabled Safety Features

The fatal collision involving an autonomous Uber vehicle and a pedestrian in Tempe, Arizona, serves as the premier case study for examining distributed causation, automation bias, and the catastrophic failure of system safety engineering. The vehicle was operating in full autonomous mode, supervised by a human safety driver. Forensic analysis of the telemetry data by the National Transportation Safety Board revealed a devastating timeline: the perception algorithms detected the pedestrian 5.6 seconds before the collision, initially failing to classify the object correctly, shifting classifications rapidly between an unknown object, a vehicle, and a bicycle. Despite the system definitively determining that emergency braking was required 1.3 seconds prior to impact, no deceleration occurred.

The engineering autopsy exposes a sequence of intentional design trade-offs that prioritized subjective performance metrics over absolute safety. The development team deliberately disabled the vehicle's factory-integrated emergency braking system specifically to eliminate "phantom braking," thereby smoothing the ride experience for the passenger. This configuration shifted the entirety of the emergency fail-safe burden onto the human safety driver.

The psychological failure inherent in this design is profound. Relying on a "human-in-the-loop" as a primary safety mechanism ignores decades of human factors engineering research regarding automation bias and vigilance decrements. When a machine performs flawlessly for extended durations, human operators inevitably suffer a degradation in situational awareness. In this instance, the driver was consuming streaming media and only registered the impending collision 0.5 seconds before impact, an intervention window that is physiologically impossible to navigate successfully.

Legally, this case highlights the immense difficulty in assigning liability across a distributed chain of actors. While the safety driver faced criminal charges for negligent

homicide due to her distraction, the deeper systemic negligence belonged to the corporate engineering team that orchestrated the removal of critical safety redundancies. The case underscores the critical necessity of teaching future engineers that human oversight cannot be utilized as a mere legal shield or an architectural crutch for imperfect algorithms. True system safety requires redundant, independent hardware fail-safes that operate autonomously of both the experimental primary software and the fallible human observer [25].

5 Pedagogical Implementation

To successfully integrate these complex legal and ethical realities into existing engineering programs, universities must move beyond theoretical lectures and implement structured, experiential methodologies. Adopting holistic frameworks—specifically the Four-Pillar Model consisting of Foundational Knowledge, Critical Appraisal, Participatory Design, and Pedagogical Integration—ensures a comprehensive educational transformation that deeply embeds contractual obligations and digital ethics into the core IT curriculum.

5.1 The "Legal Autopsy" Methodology

The "Legal Autopsy" serves as the foundational pedagogical tool for bridging the epistemological gap between abstract jurisprudence and concrete software engineering [26]. This framework requires students to systematically reverse-engineer legal failures to identify actionable technical flaws. The implementation follows a rigorous, sequential process designed to build analytical capabilities:

First, educators facilitate Case Selection, identifying a highly relevant, documented legal judgment involving algorithmic failure, such as the Air Canada misrepresentation or the Uber fatality. Second, students engage in Legal Fact Analysis, utilizing doctrinal legal analysis to thoroughly review the court's judgment, understanding the exact nature of the harm caused, the assignment of liability, and the specific legal doctrines applied, such as strict liability or duty of care. Third, the cohort conducts Technical Reconstruction, wherein engineering students map out the probable software architecture, decision trees, and data pipelines that powered the failed system, attempting to recreate the conditions of the incident.

Fourth, the process demands a rigorous Root Cause Analysis. Students must isolate the exact engineering decision, configuration parameter, or data governance failure that directly contributed to the legal breach. They must differentiate between an unforeseeable edge-case anomaly and systematic design negligence. Finally, the students execute a Redesign phase, proposing concrete, technically viable engineering solutions—such as implementing deterministic overrides, enhancing the training data representation, or integrating specific Value-Based Engineering protocols—that would have proactively prevented the legal liability. This continuous cycle of doctrinal analysis combined with technical scenario modeling drastically enhances students' critical thinking and problem-solving skills, grounding abstract ethics in tangible code.

5.2 Interdisciplinary Clinics

The successful deployment of Interdisciplinary Clinics relies on aggressive institutional support and structured faculty development aligned with Quality 4.0 dimensions, which emphasize connectivity, collaboration, and analytics to elevate educational performance. A scoping review of artificial intelligence in legal education highlights that combining theoretical legal instruction with practical, technological application is essential for anticipating the risks and effects of algorithmic implementation in complex public and private environments. These clinics must transcend traditional classroom boundaries by immersing students in high-stakes, real-world regulatory navigation.

A standard clinical model involves assembling mixed cohorts of computer science and law students, supervised jointly by faculties from both disciplines. These teams act as independent compliance consultants for external stakeholders. For instance, a university research hospital developing a new algorithmic triage system would serve as the client. The law students are responsible for conducting a comprehensive regulatory analysis, determining the system's risk classification under the EU AI Act, identifying General Data Protection Regulation vulnerabilities, and mapping the requisite transparency obligations.

Concurrently, the engineering students subject the triage model to rigorous adversarial testing, evaluating the model's mathematical fairness, robustness, and resistance to prompt injection or data drift. They analyze the explainability parameters of the outputs, ensuring the system does not operate as an impenetrable black box. The culmination of the clinical engagement is a unified impact assessment report that synthesizes legal requirements with technical realities, offering the client specific architectural modifications required to achieve legal compliance [27]. This experiential methodology heavily reinforces the reality that technological innovation and legal compliance are not mutually exclusive, but deeply interdependent.

5.3 Micro-credentials

The implementation of a micro-credentialing ecosystem must be highly modular, allowing for continuous integration of emerging regulatory standards. Universities should structure these offerings as short, intensive modules carrying continuing education units, directly targeting the most pressing industry knowledge gaps [28].

The curriculum design for these credentials must prioritize practical application over theoretical discussion. A module on "Algorithmic Bias Auditing" must involve hands-on interaction with biased datasets, teaching professionals how to apply statistical fairness metrics and document their mitigation strategies in formats compliant with the EU AI Act's technical documentation requirements. A credential in "NIST RMF Implementation" must walk engineers through the exact processes of mapping, measuring, and managing risks within an enterprise environment, culminating in the creation of a standardized compliance profile [29].

To maximize the utility of these programs, institutions must transition from legacy paper certificates to issuing cryptographically secure digital badges. By anchoring these

credentials on a blockchain ecosystem, universities empower learners with Self-Sovereign Identity, allowing them to instantly prove their competencies to global employers without relying on centralized verification authorities. This immutable ledger of continuous education ensures that as the legal landscape shifts, the workforce can rapidly, verifiably adapt, establishing a resilient talent pipeline capable of navigating the complex intersections of global algorithmic regulation.

6 Conclusion

The integration of rigorous legal competencies and deep algorithmic liability frameworks into the global engineering curriculum is no longer a theoretical, multidisciplinary luxury; it is an absolute professional necessity [30]. The "Black Box" era—characterized by the pervasive belief that engineers could successfully plead ignorance regarding the inner workings, societal consequences, and legal ramifications of their digital creations—has definitively ended. The rapid convergence of highly autonomous artificial intelligence agents, strict, divergent international regulatory empires, and a highly aggressive, evolving liability landscape demands the creation of an entirely new paradigm of engineering professional.

This new generation of "T-shaped" engineers must be technically brilliant, ethically grounded, and deeply legally literate. They must understand that the code they deploy directly dictates the boundaries of fundamental human rights, public safety, and corporate survival. By fully embracing structural pedagogical innovations—such as the rigorous "Legal Autopsy" framework, experiential interdisciplinary clinics, and the agile, blockchain-secured deployment of micro-credentials—academic institutions can effectively bridge the widening chasm between technological capability and legal accountability. Ultimately, embedding legal literacy directly into the foundational architecture of engineering education will ensure that the technologists building the future are fully equipped to design systems that are not only innovative and efficient but are fundamentally safe, transparent, and strictly compliant with the laws that govern modern society.

References

1. Ivković, R., Vučković, D., Ječmenić, M., Ananieva, D., Stanković, M., Petković, M., Janković, N.: The EU AI Act and Its Contractual and Educational Implications: AI, Education, and the Law of Obligations. International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE) **13**(2), 551–562 (2025) https://doi.org/10.23947/2334-8496-2025-13-2-551-562
2. Al-Dulaimi, A.O.M., Mohammed, M.A.-A.W.: Legal responsibility for errors caused by artificial intelligence (AI) in the public sector. Internetional Journal of Law and Management (2025) https://doi.org/10.1108/ijlma-08-2024-0295
3. Abbasnejad, B., Soltani, S., Taghizadeh, F., Zare, A.: Developing a multilevel framework for AI integration in technical and engineering higher education: insights from bibliometric analysis and ethnographic research. Interactive Technology and Smart Education **23**(1), 49–79 (2026) https://doi.org/10.1108/ITSE-12-2024-0314

4. Górski, Ł., Ramakrishna, S.: Right to Explanation in Large Language Models: Lessons from the European Union AI Act and General Data Protection Regulation. IT Professional 27, 34-40 (2025)
5. El-Khoury, M., Kpieleh, F., Khalil, S., Bou Abdo, J.: Ai Agents and Full Automation in Cybersecurity: Technical, Business, Ethical and Legal Considerations, 05 August 2025, PREPRINT (Version 1) https://doi.org/10.21203/rs.3.rs-7180712/v1
6. Chang, L., Zhou, D.: AI Integration in Higher Education: A Software Engineering Perspective on Readiness, Anxiety and Innovations. International Journal of Software Engineering and Knowledge Engineering **36**(09), 1341-1369 (2026) https://doi.org/10.1142/S0218194025501116
7. Quesada Páez, A.: La atribución de responsabilidad civil por actos de sistemas autónomos de inteligencia artificial: desafíos para el derecho privado. InDret 1.2026, 193-226 (2026) https://doi.org/10.31009/InDret.2026.i1.05
8. Tabacu, A.: Regulatory framework for the use of artificial intelligence in european higher education. Jurídicas CUC **21**(1), 124 -152 https://doi.org/10.17981/juridcuc.21.1.2025.07
9. Fattahi, A.: Global Fragmentation of AI Governance and Regulation, https://bisi.org.uk/reports/global-fragmentation-of-ai-governance, last accessed: 2026/04/15
10. Teichmann, F.: Risk, Reasonableness and Residual Harm under the EU AI Act: A Conceptual Framework for Proportional Ex-Ante Controls. European Journal of Risk Regulation, 1–25 (2026) https://doi.org/10.1017/err.2025.10077
11. Górski, Ł., Ramakrishna, S.: Right to Explanation in Large Language Models: Lessons from the European Union AI Act and General Data Protection Regulation. IT Professional 27, 34-40 (2025)
12. Wagner, M., Wittmann, M. -T., Borg, M.: A Critical Look at the EU AI Act's Requirements and Affected Systems – Who Must Explain What? In: 2025 IEEE 33rd International Requirements Engineering Conference Workshops (REW), pp. 494-503. Valencia (2025) https://doi.org/10.1109/REW66121.2025.00074
13. White House Releases National Policy Framework for Artificial Intelligence, https://www.jdsupra.com/legalnews/white-house-releases-national-policy-7560164/, last accessed: 2026/04/15
14. Swain, G.: Trump repeals Biden's AI oversight order, shifts focus to innovation-driven policies, https://www.cio.com/article/3806594/trump-repeals-bidens-ai-oversight-order-shifts-focus-to-innovation-driven-policies.html, last accessed: 2026/04/15
15. Legislative Update: Senate Removes Restrictions on AI Regulations from "Big Beautiful Bill", https://www.taftlaw.com/news-events/white-house-toolkit/senate-removes-restrictions-on-ai-regulations-from-president-trumps-big-beautiful-bill/, last accessed: 2026/04/15
16. Wilson, M.: The AI Signal from The World Economic Forum 2026 at Davos, https://globaladvisors.biz/2026/01/23/the-ai-signal-from-the-world-economic-forum-2026-at-davos/, last accessed: 2026/04/15
17. Quesada Páez, A.: La atribución de responsabilidad civil por actos de sistemas autónomos de inteligencia artificial: desafíos para el derecho privado. InDret 1.2026, 193-226 (2026) https://doi.org/10.31009/InDret.2026.i1.05
18. Stephen, L., Vairapriya: Infusion of AI technology in automated cars and its legal liability with special reference to India. In: Sundaram, A. (ed.) ARTIFICIAL INTELLIGENCE AND JUSTICE DELIVERY: OPPORTUNITIES AND CHALLENGES. pp. 353-367. The Thought Craft, Chennai (2024)

19. El-Khoury, M., Kpieleh, F., Khalil, S., Bou Abdo, J.: Ai Agents and Full Automation in Cybersecurity: Technical, Business, Ethical and Legal Considerations, 05 August 2025, PREPRINT (Version 1) https://doi.org/10.21203/rs.3.rs-7180712/v1
20. Wagner, M., Wittmann, M. -T., Borg, M.: A Critical Look at the EU AI Act's Requirements and Affected Systems – Who Must Explain What? In: 2025 IEEE 33rd International Requirements Engineering Conference Workshops (REW), pp. 494-503. Valencia (2025) https://doi.org/10.1109/REW66121.2025.00074
21. IEEE 7000-2021: Model Process for Addressing Ethical Concerns during System Design, https://quality.arc42.org/standards/ieee-7000, last accessed: 2026/04/15
22. Jonathan, G. M., Yalew, S. D. Gebremeskel, B. K., Watat, J. K.: AI Adoption in the Public Sector: Organizational Readiness and the Pursuit of Public Value. Complex Systems Informatics and Modeling Quarterly 45, 43-70 (2025) https://doi.org/10.7250/csimq.2025-45.03
23. Wei Yinze, X., Chen Wuy, L.: ESG Investment, AI Risk Management, and Ethical Compliance in Strengthening Financial Stability. Journal Economic Business Innovation **2**(2), 254-268 (2025) https://doi.org/10.69725/jebi.v2i2.272
24. Wagner, M., Gupta, R., Borg, M., Engström, E., Lysek, M.: AI Act High-Risk Requirements Readiness: Industrial Perspectives and Case Company Insights. In: Pfahl, D., Gonzalez Huerta, J., Klünder, J., Anwar, H. (eds.) Product-Focused Software Process Improvement. Industry-, Workshop-, and Doctoral Symposium Papers. PROFES 2024. Lecture Notes in Computer Science, 15453. Springer, Cham (2025) https://doi.org/10.1007/978-3-031-78392-0_5
25. Teichmann, F.: Risk, Reasonableness and Residual Harm under the EU AI Act: A Conceptual Framework for Proportional Ex-Ante Controls. European Journal of Risk Regulation, 1–25 (2026) https://doi.org/10.1017/err.2025.10077
26. Lorteau, S., Sarro, D.: Artificial intelligence in legal education: a scoping review. The Law Teacher **60**(1), 138–162 (2026) https://doi.org/10.1080/03069400.2025.2592440
27. Wagner, M., Gupta, R., Borg, M., Engström, E., Lysek, M.: AI Act High-Risk Requirements Readiness: Industrial Perspectives and Case Company Insights. In: Pfahl, D., Gonzalez Huerta, J., Klünder, J., Anwar, H. (eds.) Product-Focused Software Process Improvement. Industry-, Workshop-, and Doctoral Symposium Papers. PROFES 2024. Lecture Notes in Computer Science, 15453. Springer, Cham (2025) https://doi.org/10.1007/978-3-031-78392-0_5
28. Yıldırım, H., Kır, Şeyda, Bozkurt, A.: Towards a Future of Qualifications: A Systematic Review of Micro-Credentials. Online Learning **29**(4), 535–559 (2025) https://doi.org/10.24059/olj.v29i4.4969
29. Rullyana, G., Siregar, E., Kustandi, C.: Research Trends on Micro-Credentials in Higher Education: A Bibliometric Analysis Using Scopus and WoS Databases. Journal of Learning for Development **12**(3), 484–500 (2025) https://doi.org/10.56059/jl4d.v12i3.1950
30. AI Readiness: Building the Bridge from Higher Education to Work, https://www.pearson.com/content/dam/global-store/global/resources/ai-readiness/AI-Readiness-Report-2026.pdf, last accessed: 2026/04/15

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

