Against AI Exceptionalism: Traditional Product Liability Law as Sufficient Framework for AI Systems

Hanting Fan

I. Introduction

The AI industry has contended that neural network opacity renders traditional tort doctrine inadequate for technologies whose decision-making processes resist explanation. This AI exceptionalism narrative has influenced both litigation strategy and regulatory policy in the U.S., EU and China, with legislators and industry advocates calling for liability frameworks specifically tailored for AI. This paper argues that traditional product liability laws provide sufficient standards for AI products, as demonstrated by the recent Benavides v. Tesla verdict in the U.S., where the court applied conventional design defect and failure-to-warn doctrines to Tesla's Autopilot system. The issue lies with manufacturer non-compliance rather than legal inadequacy. Traditionally, courts in tort cases focused on evidence of safety testing, documentation of system limitations, and implementation of fail-safes. Tort law has governed complex products from pharmaceuticals to aircraft for decades, adapting to technological innovation since the Industrial Revolution without fundamental reconstruction.

II. Technical Challenges in AI Product Liability

The deep learning neural network creates faster updates of new technology and the "black box" issue that tends to avoid regulatory scrutiny and tort claims. For example, Tesla's FSD systems receive software updates wirelessly (Over-the-Air, OTA), which allows manufacturers to fix bugs remotely, improve performance, and update algorithms without physically recalling vehicles to a service center. Each OTA update would fundamentally change how the vehicle behaves because it modifies perception logic, i.e., how the vehicle "sees" its environment, and driving behavior patterns. Each update may adjust neural network parameter weights, which are the mathematical values that determine how the AI processes information, as well as perception thresholds, which are the criteria the system uses to classify objects, such as pedestrians and obstacles.

According to the current regulatory framework in the U.S., it permits manufacturers to deploy successive iterations of autonomous driving algorithms through OTA distribution mechanisms without mandatory pre-deployment submission for validation protocols or safety assessment to the authority. This constitutes a substantive departure from the established safety protocol governing automobiles for decades. In the EA22-002 investigation report conducted by the U.S. National Highway Traffic Safety Administration (NHTSA), there were instances where safety functions were temporarily degraded after updates. This means that in some cases, the new software would make certain safety features work worse than they did before, at least for a period of time.

The "black box" issue is a common challenge in litigation involving complex AI products. Modern deep learning models, like those used in Tesla's Autopilot, involve more than 48 interconnected neural networks which are trained on millions of miles of real-world driving data.

Even the creators would not be able to explain how the networks interact, weigh inputs, and arrive at a decision. The decision process emerges from billions of calculations across layers of the networks, rendering it difficult to provide a specific, step-by-step human-readable justification for every output. This opacity presents a major legal hurdle. Courts are built on centuries of rules for evidence discovery and clear accountability. In the past, the "black box" nature of AI and difficulties in accessing data made it difficult for plaintiffs to prove a specific defect. For years, Tesla sought to exclude the plaintiffs' experts by filing a Daubert motion, arguing that the plaintiff's experts' opinions were "speculative, unreliable, and not the product of reliable methodology". The essence of this legal maneuver was that without a full understanding of the complex, "black box" neural network, external experts could not form admissible, reliable opinions about the system's performance at the time of the crash.

III. Benavides v. Tesla: Traditional Tort Law in Action

On August 1, 2025, a Florida jury in Benavides v. Tesla ordered Tesla to pay a $243 million verdict in a 2019 wrongful death and injury case involving its Autopilot system, which included $43 million in compensatory damages and $200 million in punitive damages, marking a significant shift. It is the first time a jury found Tesla's Autopilot system defective and a contributing factor in a fatal crash. In the wake of the Benavides verdict, Tesla settled two other 2019 fatal crash lawsuits in California that were scheduled for trial, likely to avoid further public jury rulings. In Benavides, a Tesla Model S operating on Autopilot failed to brake while approaching a T-intersection with flashing lights, striking a parked truck and resulting in one death and severe injuries. The court granted plaintiffs' product liability claims alleging design defect and failure to warn. The jury found that there is sufficient evidence proving that Autopilot defects substantially contributed to the crash, thus making Tesla partially liable. The design defect claim succeeded without requiring Tesla to explain neural network data processing or decision-making. The plaintiffs successfully argued that Tesla's system was unreasonably dangerous due to specific design choices and marketing practices, which collectively led to a "foreseeable misuse" that the company should have prevented. The court applied conventional risk-utility analysis, examining safer alternative design feasibility and whether Tesla adequately restricted Autopilot to its operational design domain.

The court found that Tesla's driver-monitoring system, which primarily relied on steering wheel torque, was insufficient to ensure the driver remained attentive. Before the crash, Tesla's sensors detected the obstacles but the system did not brake or issue a critical warning. Moreover, the Autopilot system was defective because it could be activated on roads outside its intended operational design domain, e.g., at a T-intersection where the crash occurred, thus failing the "consumer expectations" test for safety. Tesla's marketing, including CEO Elon Musk's statements, created a false impression that "Autopilot" and "Full Self-Driving" (FSD) were more autonomous and safer than their actual Level 2 capabilities allowed, which showed that Tesla consciously disregarded known safety risks. Thus, there is a known and anticipated risk that drivers would become complacent and distracted given the product's design and marketing, i.e., foreseeable use.

In this case, instead of requiring neural network architecture understanding, the court focused on manufacturer conduct: whether Tesla defined operational domains, restricted use accordingly, and designed fail-safes for predictable misuse—all engineering choices within corporate control. In terms of the failure to warn claim, Tesla could have implemented adequate warnings through prominent in-vehicle alerts during extra-domain operation, clear dashboard displays showing system limitations, mandatory use restriction acknowledgments, or audible inattention warnings. The company's failure to employ these measures constituted inadequate warning under traditional doctrine.

The Benavides case clarifies obligations of AI system manufacturers under product liability law without neural network internal examination or AI-specific legal tests: diligence evidence in design, testing, and warning—obligations identical to those governing pharmaceuticals, medical devices, and other complex products with opaque mechanisms. The opacity defense fails through its explainability-accountability conflation—concepts tort law has always distinguished.

IV. AI-Specific Liability Regimes: An Unnecessary Departure from Established Doctrine

In 2024, the European Union has responded to these challenges with the PLD (Directive (EU) 2024/2853 on Liability for Defective Products) and the AI Act (Regulation (EU) 2024/1689), while the proposal for an AI Liability Directive has been withdrawn in February 2025. The EU's PLD imposes strict liability on manufacturers for harm caused by defective AI, regardless of fault. For example, the revised PLD explicitly expands the definition of "product" to include standalone software, AI systems, and digital components, making their developers and manufacturers liable for defects. Thus, a claimant only needs to prove that a product was defective, that they suffered damage (which now includes medically recognized psychological harm and data loss), and the causal link between the defect and the damage. The assessment of defectiveness is modernized to account for the unique nature of AI. A product can be deemed defective if:

1. It fails to meet mandatory safety requirements set out in EU law, including those in the AI Act.
2. The defect arises from insufficient software updates, a lack of necessary security patches, or cybersecurity vulnerabilities.
3. Unexpected behavior occurs due to the AI's ability to continuously learn or acquire new features after being placed on the market, as manufacturers remain liable for such outcomes.

To tackle the "black box" problem and information asymmetry, the PLD makes it easier for consumers to prove their claims in complex cases:

1. Disclosure Orders: National courts can order manufacturers to disclose relevant evidence if a claimant presents a "plausible claim" for damages. Courts must still protect trade secrets and confidential information.
2. Presumptions of Defectiveness/Causation: The defectiveness of the product is presumed if a defendant fails to disclose requested evidence, if the product doesn't comply with safety requirements, or if the damage was caused by an "obvious malfunction" under normal circumstances.

Although the PLD entered into force on December 9, 2024, the directive is not yet fully applicable in EU member states. EU Member States have a period of two years to transpose the directive's provisions into their respective national laws. Furthermore, the new rules will only apply to software and AI systems that are placed on the market or put into service after the transposition deadline of December 9, 2026. Products already on the market before December 9, 2026, will continue to be governed by the previous 1985 Product Liability Directive.

The proposed AI LEAD Act (S.2937, "Aligning Incentives for Leadership, Excellence, and Advancement in Development Act"), introduced in the U.S. Senate in September 2025 by Senators Durbin and Hawley, aims to address product liability for AI products by creating a new federal product liability framework specifically for AI systems. The Act classifies AI systems as "products," making them subject to legal scrutiny similar to physical goods. It establishes a new federal right of action allowing individuals, groups, and state or federal attorneys to sue for harms caused by an AI system in federal court. Developers could be held liable under four product liability theories: defective design, failure to warn, breach of express warranty, and strict liability for unreasonably dangerous products. Those deploying AI systems might also be liable if they significantly alter or intentionally misuse the system, though they could seek dismissal if the developer is solvent and available. The bill prohibits developers from waiving rights or unreasonably limiting liability in user agreements. Foreign AI developers would need to designate a U.S.-based agent for legal service and be listed in a public registry to deploy products in the U.S. The legislation would apply to lawsuits filed after its enactment, regardless of when the harm occurred. The goal is to encourage AI companies to prioritize safety in design and development through clear rules and accountability. The bill is pending in the U.S. Senate committee.

AI-specific liability regimes suffer from four fundamental flaws that undermine accountability. By creating specialized frameworks, legislators validate the claim that AI products require different legal treatment. Throughout Benavides, Tesla argued neural network opacity rendered traditional analysis inapplicable. However, the jury rejected this, applying conventional design defect doctrine. New AI-specific legislation resurrects the exceptionalism defense that Benavides refuted, giving manufacturers renewed grounds to argue traditional standards cannot govern their products. Furthermore, the overlapping regimes would create strategic characterization opportunities, i.e., manufacturers can emphasize or downplay AI components depending on which regime proves less stringent. For example, a semi-autonomous vehicle combining mechanical and AI systems faces uncertain treatment, increasing litigation costs and enabling forum-shopping. Moreover, the PLD's two-year transposition period and prospective-only application create an accountability vacuum. Products deployed before December 9, 2026, escape new requirements regardless of AI sophistication.

V. Conclusion

The Benavides verdict demonstrates what legislators drafting AI-specific frameworks have overlooked: traditional product liability doctrine already provides the necessary tools for AI accountability. The court's application of century-old design defect and failure-to-warn principles to Tesla's Autopilot system required no specialized legal innovation, no neural network transparency, and no legislative intervention. Instead, it focused on what tort law has always demanded—evidence of reasonable design choices, adequate warnings, and fail-safe implementations. The rush to create AI-specific liability regimes rests on a false premise. Neural network opacity does not render traditional legal analysis inadequate; it simply shifts the evidentiary focus from internal algorithmic processes to external manufacturer conduct. Courts have governed complex products with opaque mechanisms for decades, from pharmaceuticals to medical devices, without abandoning foundational principles. AI products are not exceptional. The path forward requires rejecting AI exceptionalism, not codifying it through fragmented, delayed, and ultimately counterproductive specialized frameworks.

REFERENCES

1. Abbott, R. (2020). The reasonable robot: Artificial intelligence and the law. Cambridge University Press.
2. Buiten, M. C., De Streel, A., & Peitz, M. (2023). The law and economics of AI liability. Computer Law & Security Review, 48, Article 105794.
3. Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. UC Davis Law Review, 51(2), 399–435.
4. European Commission Directorate-General for Justice and Consumers. (2024). Directive (EU) 2024/2853 on liability for defective products. Official Journal of the European Union.
5. European Parliament Research Service. (2024). Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence. European Union.
6. Geistfeld, M. (2025). Product liability law in the age of AI (3rd ed.). Aspen Publishing.
7. Hacker, P. (2023). The European AI liability directives: Critique of a half-hearted approach and lessons for the future. Computer Law & Security Review, 51, Article 105871.
8. Sharkey, C. M. (2024). A products liability framework for AI. Columbia Science and Technology Law Review, 25(2), 275–334.
9. Smith, B. W. (2017). Automated driving and product liability. Michigan State Law Review, 2017(1), 1–74.
10. Wischmeyer, T., & Rademacher, T. (Eds.). (2020). Regulating artificial intelligence. Springer.