Genbounty

ISITC EUROPE CIC

WHO NEEDS AI PRODUCT SAFETY?

An ISITC Europe/Genbounty White Paper

Abstract

Deployers of AI systems and developers of AI applications who intend to conduct business within the EU are now classed as manufacturers of products, with all the associated liability for consumer protection

Robert Morel

Contents

1. Executive Summary

Artificial intelligence is no longer just code, it is now legally recognized as a product in the European Union, subject to the same rigorous safety standards as physical goods. The EU AI Act and the revised Product Liability Directive have redefined the landscape: any organization deploying or selling AI systems in the EU is now considered a manufacturer, directly liable for consumer protection.

This shift is driven by the urgent need to protect consumers from tangible harms caused by unsafe or deceptive AI products. These risks include physical danger, psychological harm, fraud, illegal discrimination, violations of fundamental rights, and manipulative techniques. The EU AI Act enforces strict obligations for high-risk AI systems, mandating robust safety assessments, transparency, human oversight, and clear documentation. The Product Liability Directive complements this by establishing a strict liability regime, victims need only prove that an AI product was defective and caused harm, not that the developer was negligent.

For businesses, this means AI product safety is no longer a voluntary ethical guideline but a mandatory engineering discipline. Compliance is now a prerequisite for market access and legal protection. Organizations must adopt end-to-end processes for AI safety, from initial testing and documentation to post-market monitoring and accreditation.

Internationally, the EU's comprehensive approach is setting a global benchmark, even as other regions like the United States pursue more voluntary, market-driven strategies. Industry standards such as ISO/IEC 42001:2023 are emerging to guide best practices worldwide.

In summary, the era of treating AI safety as an abstract concept is over. The new legal paradigm demands that organizations prioritize consumer-first product safety, implement robust governance, and ensure compliance at every stage of the AI lifecycle. Those who fail to adapt risk will suffer not only regulatory penalties but also significant financial and reputational harm.

2. Key Risks for Capital Markets Businesses

2.1. Legal and Regulatory Liability

With the EU AI Act and the revised Product Liability Directive, AI systems, including those used in trading, risk management, compliance, and client services, are now legally classified as products. This means capital markets firms deploying AI in the EU are directly liable for any harm caused by defective or non-compliant AI systems. The strict liability regime means that victims do not need to prove negligence, only that the AI product was defective and caused harm.

2.2. Financial and Litigation Exposure

Defective AI systems can result in significant financial losses, not only from direct operational failures (such as erroneous trades or risk miscalculations) but also from lawsuits and compensation claims. The new legal framework lowers the burden of proof for claimants and empowers courts to demand technical evidence from firms, increasing the risk of costly litigation.

2.3. Reputational Damage

Incidents involving AI-driven errors, such as biased credit decisions, market manipulation, or failure to detect fraud, can quickly erode trust among clients, investors, and regulators. In capital markets, where reputation is critical, such events can have long-lasting negative effects.

2.4. Operational Risks and Compliance Failures

AI systems in capital markets must be robust, reliable, and fair. Failures in these areas can lead to market disruptions, regulatory sanctions, and loss of business. The EU AI Act mandates rigorous safety assessments, transparency, and human oversight for high-risk systems, making compliance a complex and ongoing operational challenge.

2.5. Consumer Duty and Market Harm

Unsafe or deceptive AI products can cause direct harm to consumers and market participants, including:

- Financial loss due to algorithmic errors or manipulation
- Unfair or discriminatory outcomes in lending, onboarding, or trading
- Violation of privacy or fundamental rights through unlawful surveillance or data misuse

In summary, capital markets firms face a new era of accountability for AI systems. The risks span legal, financial, operational, and reputational domains, making robust AI product safety and compliance not just a regulatory requirement but a business imperative.

3. AI Product Safety Components

3.1 What Makes a Safe AI Product?

A safe AI product is a system that is demonstrably and verifiably sound across several key characteristics. These components are the building blocks of trustworthy and responsible AI.

- **Validity and Reliability:** The system must function as specified and do so consistently. Validity asks if the AI system is accurate in its conclusions. Reliability asks if it produces those conclusions consistently over time and under a range of conditions.
- **Robustness:** The system must maintain its specified function and performance, even when faced with unexpected inputs, edge cases, or adversarial attempts to make it fail.
- **Fairness and Bias Mitigation:** The system must not produce inequitable, discriminatory, or unjust outcomes. As defined by the National Institute of Standards and Technology (NIST), this is a foundational building block of trustworthiness and an essential component of safety and reliability.
- **Accountability:** There must be unambiguous ownership over AI systems, their impacts and resulting outputs across the AI lifecycle. This is a critical governance component, requiring that design decisions such as the trade-off between a model's performance and its explainability are documented, along with who made the decision and when. Accountability must be both proactive - preventing harms before they occur, and reactive - responding proportionately to incidents.

3.2 Consumer-first Product Safety

The immediate answer to "who needs AI product safety" is Consumers, meaning your customers. This need is not for protection against speculative future harms or existential risk, but from clear and present implications of real-world consumer harm. This means AI product safety should follow a consumer-safety first approach.

The EU AI Act has become a primary enforcer in this domain for who intend to conduct business within the EU, defining AI safety as a framework for consumer protection and the protection of consumer fundamental rights to expect a safe product.

Consumers are encountering AI systems, often without their knowledge, in everyday products. These include customer service chatbots, educational tools, social media recommendation feeds, voice assistants, recommendation engines, and smart devices. The harms that the European Commission has identified as resulting from unsafe or deceptive AI products are tangible:

- **Physical Harm or Danger:** Medical misdiagnosis, autonomous vehicle malfunctions, unsafe industrial controls
- **Psychological Harm:** Suicide encouragement, harassment, traumatic content, targeting vulnerable individuals.
- **Fraud and Impersonation:** AI-generated deepfakes and voice cloning tools are enabling fraud and sophisticated impersonation scams at scale.
- **Illegal Discrimination:** AI tools can perpetuate illegal discrimination. This is most dangerous when AI products make decisions in high-risk contexts, determining whether a consumer gets medical help, a place to live, a job, or a loan.
- **Fundamental Rights Violations:** Unlawful surveillance, censorship, denial of benefits or essential services.
- **Harm to Vulnerable Groups:** Unsafe child interactions, exploitative targeting of minors or elderly.
- **Manipulative Techniques:** Subliminal nudging, deceptive conversations, undue influence causing harm.

4. The Business Case for AI Product Safety

The risk of consumer harm presents an immediate need for organizations that develop and deploy these technologies to meet AI product safety requirements. Firms now face serious financial, legal, and reputational liabilities when deploying AI products. The business case for AI product safety is to achieve compliance, for the protection of consumers, and the prevention of associated litigation.

4.1 AI as a Product

Historically, Europe centric legal frameworks have struggled to apply product liability law to "intangible" software. Software failures were often treated as breaches of contract or service, not as product defects. But the European Union's updated Product Liability Directive (PLD) fundamentally changes this. It recognizes that AI is not just code, it is a product integrated into daily life that can cause tangible physical, psychological, and financial harm.

Instead of waiting for real world litigation to test legal theories, the EU has proactively legislated this reality. The new Product Liability Directive is the cornerstone of this strategy. It explicitly expands the definition of "product" to include software, digital manufacturing files, and standalone AI systems.

This change represents a fundamental legal shift. It means an AI chatbot, algorithm, or embedded system is treated just like a faulty car or a mislabelled pharmaceutical. If the AI is defective and causes harm, its manufacturer (the developer or provider) can be held liable. The AI developer is now the manufacturer of a product, not a neutral intermediary. This makes organizations intending to sell AI products within the EU are directly liable for the product's defects, which can include flawed generated content, biased decision making, or unsafe autonomous actions.

This legal shift for providers of AI systems within the EU is built on a set of interconnected regulations that establish what a defect is and how to prove it:

The EU AI Act: This regulation defines the "standard of care." For high-risk AI systems, it mandates rigorous safety assessments, technical documentation, human oversight, and clear warnings. A failure to meet these requirements can be used as direct evidence that the AI system was defective.

The Product Liability Directive (PLD): This provides the "cause of action." It establishes a strict liability regime. This means a victim does not need to prove the developer was negligent. They only need to prove the AI product was defective (e.g., by not complying with the AI Act) and that the defect caused their harm.

4.2 The EU AI Act (AIA)

The European Union has taken a comprehensive and mandatory approach to AI product safety. The EU AI Act, which came into force in 2024, is the first comprehensive legislation in the world regulating artificial intelligence. It establishes a clear, risk-based framework:

Unacceptable Risk (Banned): This category includes AI systems that pose a clear threat to fundamental rights. It bans systems for cognitive behavioural manipulation, social scoring by governments, and most real-time biometric identification in public spaces.

High-Risk Systems (Strictly Regulated): This is the core of the Act. An AI system is classified as high-risk if it falls into one of two categories:

1. AI is used as a safety component in products already covered by EU product safety laws.
2. AI systems used in specific sensitive areas (e.g., management of critical infrastructure, financial decision making, education and vocational training, employment and worker management, law enforcement, and the public sector).

Obligations for High-Risk Systems: Providers (developers) of these systems are subject to strict, mandatory obligations before the product can be put on the market. These include conducting adequate risk assessments, using high-quality training data to minimize bias, logging all activity for traceability, providing clear instructions for use, and ensuring appropriate human oversight, robustness, and cybersecurity.

General-Purpose AI (GPAI): All GPAI models, including systems like ChatGPT, must adhere to fundamental transparency obligations, such as documenting the model, providing clear instructions for use, and

implementing policies to respect copyright law during training. Content generated by any GPAI must be clearly disclosed as AI-generated.

While the Act is in force, its application is staggered from 2025 Unacceptable risk, 2026, high risk, through 2027, full rollout.

4.3 The Product Liability Directive (PLD)

The updated Product Liability Directive (PLD) provides consumer protection through enforcement and compensation. Adopted in 2024, this directive overhauls the EU's 40-year-old liability rules specifically to address the digital age and AI. It is the legal mechanism that allows citizens to seek compensation for harm.

- **Expanded Definition of product:** This is the PLD's most critical update. It explicitly clarifies that "products" include software, AI systems, and digital manufacturing files. This change definitively ends the legal debate over whether intangible software could be subject to product liability, making AI developers manufacturers under the law.
- **Works with the AI Act:** The PLD is designed to work alongside the AIA. A key way to prove an AI product was "defective" is to show that it failed to comply with the mandatory safety requirements of the AI Act. If a high-risk AI system causes harm and it is found to be non-compliant with the AI Act's rules, it can be presumed defective.
- **Easing the Burden of Proof:** The directive tackles the AI black box problem. It gives national courts the power to order a developer to disclose technical information and evidence. Furthermore, the law establishes a "presumption of defect" if the manufacturer fails to disclose this court-ordered information, or if it is proven they violated the AI Act, thereby shifting the burden of proof.
- **Strict Liability (No-Fault):** The PLD maintains its core principle of "strict liability." A victim does not need to prove the developer was negligent or intended to cause harm. They only need to prove the AI product was defective and that the defect caused their damage to a person or their property.

4.4 International and Industry Standards

The United States has, to date, pursued a different path, prioritizing a voluntary, market-driven, and non-legislative approach. The goal is to foster innovation while encouraging safety, rather than mandating it. Unlike the European Union's AI Act, which provides a legally binding framework, the U.S. adopts a decentralized, sector-specific regulatory strategy, primarily driven by voluntary commitments from private companies and guided by federal agencies.

Filling the gap between the EU and US models is a rapidly growing body of international law and industry standards. These are not currently legally mandatory, but they are forming a global consensus on best practices.

- **Global Organizations:** The OECD, UNESCO, and the G7 are all driving initiatives to create interoperable AI governance frameworks, often cantered on principles like human rights, fairness, and accountability.
- **ISO/IEC 42001:2023:** This important, practical standard is a global, auditable benchmark for establishing an AI Management System (AIMS) within an organization. It provides a structured framework for managing AI development, deployment, compliance, and trust-building.

This global landscape is defined by the conflict between two regulatory philosophies. The Brussels Effect is the theory that the EU AI Act, by being the most comprehensive and restrictive, will become the de facto global standard, as multinational companies will find it easier to apply the strictest rule everywhere. However, this is being met by a powerful counterforce, as U.S. businesses and political actors resist, arguing that mandatory compliance undermines the benefits of AI and stifles innovation.

5. Conclusions

The time for treating AI safety as an abstract ethical guideline or a voluntary corporate initiative is over. The central argument is that AI Product Safety is now a legally binding engineering and compliance discipline with urgent implications for organizations to develop policies and governance to manage AI product risk.

For consumers, this represents a critical and overdue response to the tangible, real-world harms - from systemic bias in finance and employment to physical danger and psychological manipulation that unsafe AI products present.

For businesses, the realization that AI is classed as a product should be a swift wake up call. AI Product safety has now shifted from an ethical consideration to a non-negotiable prerequisite for market access and legal compliance.

The European Union has acted as the primary catalyst for this change. By strategically combining the EU AI Act as the new “standard of care” with the Product Liability Directive (PLD) as the cause of action, regulators have systematically dismantled the legal ambiguity that once shielded software. AI is no longer an intangible service but a product, and its developers are manufacturers subject to the same strict liability as those who build physical products.

This shift from an ethical “should" to a legal “must” raises the single most important question for businesses: What is the practical path to compliance? Organizations now require a structured, end-to-end process to manage this new liability. This process must be a clear pathway that integrates safety from initial testing and compliance documentation all the way to post-market monitoring and accreditation.

ISITC Europe in partnership with Genbounty offers member firms within the Capital Markets an independent assessment and accreditation to assist in compliance with the aforementioned directives. Further information can be found on the ISITC Europe AI Forum page.

References

AI Alignment Forum (2025) AI Safety Strategies Landscape. Available at: https://www.alignmentforum.org/posts/RzsXRbk2ETNqjhsma/ai-safety-strategies-landscape (Accessed: 22 October 2025).

arXiv (2025b) What Is AI Safety? What Do We Want It to Be? Available at: https://arxiv.org/html/2505.02313v1 (Accessed: 4 November 2025).

Davtyan, Tatevik, The U.S. Approach to AI Regulation: Federal Laws, Policies, and Strategies Explained (September 09, 2024). Available at SSRN: https://ssrn.com/abstract=4954290 or http://dx.doi.org/10.2139/ssrn.4954290

DataSunrise (2025) AI Safety vs AI Security: Trustworthy Artificial Intelligence. Available at: https://www.datasunrise.com/knowledge-center/ai-security/ai-safety-vs-ai-security/ (Accessed: 7 November 2025).

Diligent (2025) NIST AI Risk Management Framework: A simple guide to smarter AI .... Available at: https://www.diligent.com/resources/blog/nist-ai-risk-management-framework (Accessed: 12 November 2025).

EU Artificial Intelligence Act (2024) High-level summary of the AI Act. Available at: https://artificialintelligenceact.eu/high-level-summary/ (Accessed: 28 October 2025).

IBM (2025a) What is AI Ethics? Available at: https://www.ibm.com/think/topics/ai-ethics (Accessed: 2 November 2025).

IBM (2025b) What Is AI Safety? Available at: https://www.ibm.com/think/topics/ai-safety (Accessed: 25 October 2025).

Lewis Silkin LLP (2023) Discrimination and bias in AI recruitment: a case study. Available at: https://www.lewissilkin.com/en/insights/2023/10/31/discrimination-and-bias-in-ai-recruitment-a-case-study (Accessed: 11 November 2025).

OneTrust (2023) Navigating the NIST AI Risk Management Framework with confidence | Blog. Available at: https://www.onetrust.com/blog/navigating-the-nist-ai-risk-management-framework-with-confidence/ (Accessed: 16 October 2025).

Palo Alto Networks (2023) NIST AI Risk Management Framework (AI RMF). Available at: https://www.paloaltonetworks.com/cyberpedia/nist-ai-risk-management-framework (Accessed: 8 November 2025).

Squire Patton Boggs (2023) Reconciling Artificial Intelligence (AI) With Product Safety Laws. [pdf] Available at: https://www.squirepattonboggs.com/-/media/files/insights/publications/2023/12/reconciling-artificial-intelligence-ai-with-product-safety-laws/reconciling-artificial-intelligence-ai-with-product-safety-laws.pdf?rev=19e571d0581c4e50a91d2fb6a60c3b22&hash=16F89B7FF4AAC052A3446499F9518F6F (Accessed: 21 October 2025).

Stanford AI Alignment (2025) SAIA – Stanford AI Alignment. Available at: https://stanfordaialignment.org/ (Accessed: 5 November 2025).

The Guardian (2025) EU could water down AI Act amid pressure from Trump and big tech. Available at: https://www.theguardian.com/world/2025/nov/07/european-commission-ai-artificial-intelligence-act-trump-administration-tech-business (Accessed: 9 November 2025).

UNESCO (2021) Ethics of Artificial Intelligence. Available at: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics (Accessed: 27 October 2025).

Wikipedia (2025) AI safety. Available at: https://en.wikipedia.org/wiki/AI_safety (Accessed: 31 October 2025).

About ISITC Europe CIC

Is a not-for-profit industry body promoting education, innovation, and collaboration in global Capital Markets.

About Genbounty

Genbounty is a product-level AI safety testing platform with native EU AI Act compliance built in. The platform includes consumer reporting, post-market monitoring, and Product Liability Directive assurance to protect your users, and your company.

Disclaimer