June 2026

COMMENTS IN RESPONSE to Call for Consultation on European standard AI Risk Management prEN 18228 Draft

Authors: Ahmed Nagy, Paolo Giudici

ACM Europe Technology Policy Committee

Introduction

The draft prEN 18228:2026 is a significant and constructive contribution to the European AI governance landscape. It provides a structured risk management process aligned with the EU AI Act and brings together several essential elements: life-cycle risk management, intended purpose, reasonably foreseeable misuse, risk acceptability criteria, testing, risk control, residual-risk evaluation, management review, and pre-market/post-market monitoring. Its strongest policy contribution is that it treats fundamental rights as first-class risk concerns alongside health and safety. This is important because many AI harms do not arise only from technical malfunction, but from systems that operate effectively while producing unjustified exclusion, discrimination, opacity, surveillance, or unequal access to rights and services.

Analysis of the Draft

From the perspective of a data-focused technology policy committee, however, the draft should be strengthened in four areas. First, it needs a **clearer proportionality model**. A single conceptual risk management process is useful, but implementation should differ substantially between low-risk productivity tools, middle-tier decision-support systems, and high-risk AI systems affecting rights, safety, livelihood, public services, children, vulnerable groups, or critical infrastructure. Without such tiering, the standard may overburden low-risk innovation while leaving too much discretion in high-risk settings.

Second, the draft should make **data governance more central.** AI risks are frequently rooted in data provenance, data quality, representativeness, metadata, synthetic data, lineage, lawful access, data minimization, and drift. These issues should be treated not merely as supporting

evidence for risk assessment, but as a core layer of AI risk management. The standard should therefore include clearer requirements for data documentation, data contracts, quality controls, subgroup analysis, drift monitoring, and lifecycle data accountability.

Third, the standard should **clarify what counts as sufficient objective evidence for risk** acceptability. For low-risk systems, lightweight internal assessment may be sufficient. For middle-tier systems, structured testing, bias checks, domain validation, and monitoring metrics should be expected. For high-risk systems, independent multidisciplinary review, stakeholder consultation, fundamental-rights analysis, subgroup testing, and top-management approval should be required. The draft already contains the foundations for such an approach, including objective evidence, consultation, vulnerable-group considerations, and independence, but it should make the evidence ladder more explicit.

Fourth, the draft should better **address modern AI supply chains and general-purpose AI dependencies**. Many AI systems are built on foundation models, APIs, third-party datasets, fine-tuned components, retrieval systems, and external infrastructure. Risk management must therefore include dependency mapping, model versioning, vendor controls, change management, evaluation boundaries, and fallback procedures.

ACM welcomes the draft's attempt to translate AI Act risk management into a concrete, life-cycle process. The focus on intended purpose, reasonably foreseeable misuse, hazard identification, risk estimation, testing, risk control, residual-risk evaluation, review, and pre/post-market monitoring is directionally sound.

Recommendations to the CEN/CENELEC JTC21 Committee

1. Preserve the integration of fundamental rights as a core contribution.

The draft's inclusion of fundamental rights, affected persons, vulnerable groups, consultation, and rights interference is essential. This should not be weakened in later revisions. It is one of the main features that makes the draft suitable for European AI governance rather than only technical product assurance.

2. Add an explicit proportionality model.

ACM should recommend a tiered implementation structure for low-, medium-, and high-risk AI systems. The same conceptual process can apply across tiers, but the required evidence, independence, testing, consultation, documentation, and review intensity should scale with potential harm.

3. Strengthen data governance.

The standard should include a dedicated data-risk management layer. Data quality, provenance, representativeness, lineage, synthetic data, data contracts, data minimization, drift monitoring, and access governance are not peripheral issues; they are primary sources of AI risk.

4. Clarify evidence sufficiency and independent review.

The standard should define what counts as sufficient objective evidence at different risk levels. For high-risk systems, internal evidence alone should not normally be sufficient. Independent review, multidisciplinary expertise, stakeholder consultation, and post-market feedback should be expected.

5. Clarify deployer and downstream responsibilities.

AI risk often changes in deployment. ACM should recommend clearer requirements for handover information, deployment-context reassessment, feedback loops between deployers and providers, and responsibilities when systems are integrated, fine-tuned, customized, or used beyond their original context.

6. Expand guidance for general-purpose AI dependencies.

The standard should address systems built on foundation models and third-party AI services. Providers should document model versioning, evaluation limits, dependency risks, update controls, fallback plans, and contractual assurances.

7. Make post-market monitoring more operational.

The draft should define examples of measurable post-market triggers, such as model drift, subgroup performance degradation, repeated complaints, serious near misses, cybersecurity events, excessive human override, and evidence of rights interference.

8. Provide practical annexes and templates.

To avoid compliance becoming a burden for SMEs and public institutions, the final standard should include sample risk files, low-risk templates, medium-risk checklists, high-risk evidence matrices, and examples of acceptable and unacceptable risk acceptability criteria.

9. Clarify how governance obligations scale with risk.

Clearly state that higher-risk AI systems require stronger oversight, stronger evidence, more rigorous testing, more frequent review, clearer escalation procedures, and greater independence in risk evaluation.

10. Distinguish among risk categories

Distinguish among the following categories for risk in the governance and treatment cycle: 1) Known risk with estimable probability, 2) known risk with uncertain probability, 3) emergent risk, 4) systemic or cumulative risk, and 4) near-miss risk

Reflection on the definition of risk

The draft defines risk as having two key components: **the likelihood (probability of occurrence) of a harm** and **the severity of that harm**. It also defines risk formally as the "combination of the probability of an occurrence of harm and the severity of that harm."

This is useful, but it creates a conceptual challenge for AI. In many AI systems, probability is not stable or easy to estimate. For example, the probability of harm may change because of new deployment contexts, model updates, data drift, user adaptation, adversarial use, supply-chain changes, or interaction with other AI systems. In fundamental-rights cases, the harm may be hard to quantify: a privacy interference, discriminatory exclusion, chilling effect, or denial of procedural fairness may not appear as a simple measurable event.

The draft partially recognizes this problem. It says that risk acceptability criteria must be established **even when the probability of harm cannot be quantified**. That is a strong point, but it should be made more central. For AI, the standard should distinguish among:

- **Quantifiable likelihood and harm :** We have enough evidence to estimate both likelihood and severity.
- **Quantifiable harm, qualitative probability::** We have enough evidence to estimate the harm, but the likelihhod cannot be reliably quantified, and should be estimated in a qualitative manner
- **Quantifiable likelihood, emergent risk:** We have enough evidence to estimate the harm, but the risk risk appears only after deployment, integration, scaling, or new user behavior.
- **Quantifiable likelihood, Near-miss risk: the likelihood is quantifiable, but the** harm did not occur, but the system entered a hazardous state where serious harm could reasonably have occurred.
- **Systemic risk:** Individual incidents may look minor in likelihood and severity,, but repeated patterns create serious rights, safety, or social harms.

The draft correctly defines risk in terms of probability and severity, but AI providers need more help translating this into concrete indicators. The standard says the system used for qualitative or quantitative categorization of probability and severity must be documented and justified, and that risk estimation can be qualitative and quantitative. However, it does not provide enough guidance on the distinction between qualitative and quantitative assessments. It should also provide AI-specific metric families. It could benefit from an annex listing candidate indicators for accuracy, robustness and security, fairness, explainability and human oversight, , data quality.

The draft standard treats probability as probability of harm. A model error is not automatically a harm. Conversely, harm can arise even without a model error, for example when a technically correct prediction is used in an unfair, disproportionate, or unlawful decision process. The standard should distinguish between risk drivers, which may cause riks, and actual risks, which are their material occurrence.

It is recommended that the standard distinguish between:

- **Risk driver :** Probability that a person or group is exposed to a hazardous situation, because of a risk driver, such as model error
- **Risky**: Probability that exposure actually produces health, safety, rights, financial, or social harm.
- **Actual risk:** Probability that a harmful or near-harm event occurs in operation.

Conclusions

We support the direction of prEN 18228 while recommending targeted revisions. The final standard should remain rights-preserving, evidence-based, auditable, and life-cycle-oriented, but it should become more operational, tiered, data-aware, and practical for organizations of different sizes and risk profiles. Its success will depend on whether it becomes a usable governance instrument rather than a generic compliance checklist.

References

CEN-CENELEC. (2026). prEN 18228:2026: AI risk management [Draft European Standard]. CEN-CENELEC.

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union.

Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial intelligence risk measurement. Expert Systems with Applications, 235, Article 121220.

International Organization for Standardization. (2009). ISO Guide 73:2009: Risk management vocabulary. ISO.

International Organization for Standardization. (2018). ISO 31000:2018: Risk management guidelines. ISO.

International Organization for Standardization and International Electrotechnical Commission. (2014). ISO/IEC Guide 51:2014: Safety aspects, guidelines for their inclusion in standards. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2021). ISO/IEC TR 24027:2021: Information technology, artificial intelligence, bias in AI systems and AI aided decision making. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2021). ISO/IEC TR 24029-1:2021: Artificial intelligence, assessment of the robustness of neural networks, Part 1: Overview. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2023). ISO/IEC 23894:2023: Information technology, artificial intelligence, guidance on risk management. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2023). ISO/IEC 42001:2023: Information technology, artificial intelligence, management system. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2023). ISO/IEC 25059:2023: Software engineering, systems and software Quality Requirements and Evaluation (SQuaRE), quality model for AI systems. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2024). ISO/IEC TS 12791:2024: Information technology, artificial intelligence, treatment of unwanted bias in classification and regression machine learning tasks. ISO/IEC.

International Organization for Standardization and International Electrotechnical Commission. (2025). ISO/IEC TS 6254:2025: Artificial intelligence, objectives and approaches for explainability of ML models and AI systems. ISO/IEC.

International Organization for Standardization, International Electrotechnical Commission, and Institute of Electrical and Electronics Engineers. (2022). ISO/IEC/IEEE 29119-1:2022: Software and systems engineering, software testing, Part 1: General concepts. ISO/IEC/IEEE.

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.