Nemko Digital

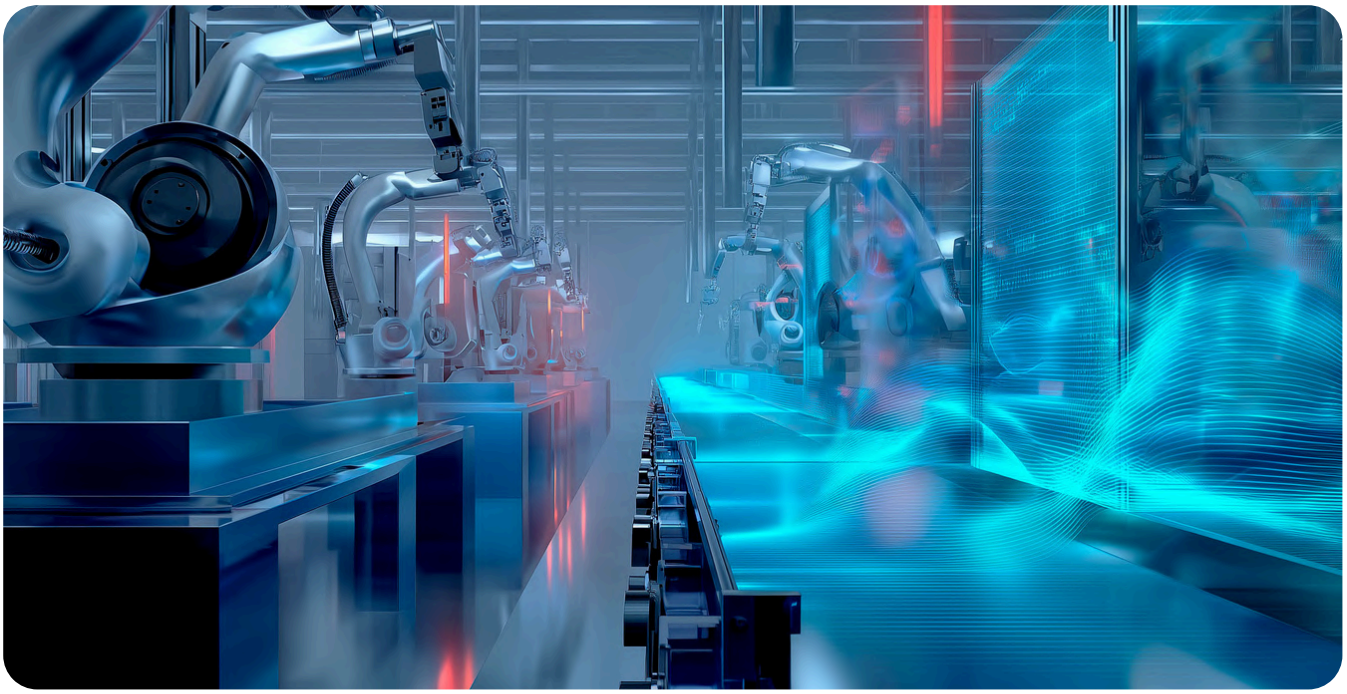
From Risk to Resilience:

Assurance-Driven AI in the Technical Industry

"AI is no longer a lab experiment; trust must be designed in from day one to make AI viable at scale in industry."

Across the technical and industrial domains, artificial intelligence is rapidly moving from proof-of-concept to production. AI is now embedded in machinery, industrial software, autonomous systems, and decision-making processes that directly affect safety, reliability, and business continuity. As this transition accelerates, the primary challenge is no longer whether AI can deliver value, but whether organizations can deploy it safely, responsibly, and with enduring trust.

In the technical industry, trust in AI is inseparable from risk management. Where AI systems influence physical processes, safety-critical functions, or compliance obligations, trust must be engineered through structure, governance, and independent assurance. This is not about slowing innovation; it is about enabling AI to scale without creating unacceptable operational, safety, or reputational risks.



AI Raises the Bar for Risk Management

Industrial organizations are deeply familiar with managing risk. For decades, safety, quality, security, and compliance have been governed through standards, management systems, and certification. AI, however, introduces a new class of complexity.

Unlike deterministic software, AI systems behave probabilistically, are heavily data-dependent, can evolve over time, and may lack explainability.

These characteristics amplify traditional risks and introduce new ones; especially when AI interacts with physical systems. As discussed in Dissecting What's Needed to Scale Agentic AI with Confidence, the challenge is not autonomy itself, but unmanaged autonomy without clear technical and organizational guardrails.

This means organizations must consciously extend their existing risk management to include AI-specific risks, and anchor these controls in a formal AI management system that ensures responsibility, oversight, and continuous control as AI systems scale and evolve.

Functional Safety: A Critical Requirement

Traditional safety frameworks are built around systems that behave in a **deterministic and predictable way**. Functional safety requirements for machinery, control systems, and industrial electronics are derived from clearly defined functions, fixed logic, and verifiable cause-and-effect relationships. In that context, safety can be specified, tested, and certified against well-understood criteria.

AI fundamentally challenges this model

One of the most persistent gaps in AI adoption today is the insufficient integration of functional safety principles in an environment where system behavior is no longer purely deterministic. AI systems introduce probabilistic decision-making, data-dependent behavior, and variability in outcomes. This represents a new perspective for an industry used to requirements that do not change depending on statistical inference or learned behavior.

NEMKO previously described this challenge explicitly as a missing safety layer when AI becomes part of physical systems.

When AI influences control decisions (safety-related functions, autonomous or semi-autonomous behaviour, human-machine interaction) it becomes part of the safety equation. Treating AI as "just software" creates fragmented accountability and hidden systemic risks.

Functional safety requires:

- Clearly defined intended use and limits
- Hazard identification and risk analysis
- Safety integrity requirements
- Verification, validation, and lifecycle control

These principles must be extended to AI-enabled systems. The aim is not to force AI into legacy frameworks, but to evolve safety engineering so AI failure modes, uncertainty, and human oversight are explicitly managed.

Addressing this gap requires extending functional safety engineering to include AI-specific behavior through systematic testing, validation, and monitoring—ensuring that probabilistic outcomes, failure modes, and human oversight are explicitly verified and kept under control throughout the system lifecycle.

Supply-Chain Responsibility: Trust Does Not Stop at the Factory Gate

One of the most underestimated aspects of AI risk lies in the AI supply chain. Very few organizations develop AI systems entirely in-house; instead, AI enters through vendors, cloud services, embedded software, or certified products. As a result, trust shifts from development alone to procurement and supplier governance.

Comparable to a Software Bill of Materials (SBOM), responsible AI deployment requires transparency into:

- Embedded models and algorithms
- Data sources and training assumptions
- Third-party dependencies
- Update, retraining, and decommissioning mechanisms

Critically, accountability for AI behaviour does not disappear when AI is purchased rather than built. Organizations remain responsible for how AI operates within their systems and products, regardless of where it originated.

Embedding AI governance into procurement therefore means requiring suppliers to demonstrate:

- Clear design intent and defined limits of use
- Structured risk assessment and mitigation
- Evidence of testing and validation
- Provisions for monitoring across the AI lifecycle

Without these controls, organizations unknowingly import risks they neither understand nor manage. In industrial contexts, trust is never based on intent alone—it is based on **demonstrable control**. For AI acquired through the supply chain, this control must be verifiable through governance, engineering evidence, and ongoing oversight. Testing, monitoring, and transparency are not separate activities, but the means by which responsibility for externally sourced AI can be proven and sustained over time.

Transparency is not a separate requirement but the outcome of effective governance and assurance. It becomes proof of control: showing that AI systems are understood, managed, and trusted, even as they grow more autonomous and complex.

From Innovation Risk to Industrial Resilience

Stepping back from the individual building blocks (governance, functional safety, supply-chain control, and assurance) a common pattern emerges: trust in industrial AI ultimately depends on knowing what is actually being trusted.

Many organizations struggle with AI not because it is inherently unsafe, but because they lack a coherent overview of how AI is used across the organization. AI often enters through multiple paths (innovation initiatives, procured software, embedded product functionality, or operational tooling) without being recognized as part of a single risk and control landscape.

Effective trust begins with asking a small number of fundamental questions:

- Which AI systems are in use, and where in the organization or value chain?
- Are they stand-alone tools, embedded in software, or integrated into physical products and systems?
- What decisions or actions do they influence (advisory, operational, or safety-related)?
- What happens when those systems fail, drift, or behave in unintended ways?

As explored earlier in Dissecting What's Needed to Scale Agentic AI with Confidence, risk increases significantly as AI moves from decision support toward partial or full autonomy. At that point, questions of oversight, responsibility, and acceptable behavior can no longer be addressed locally or informally.

Viewed holistically, AI risk in the technical industry typically spans multiple dimensions:

- **Operational risk –** incorrect outputs, degraded performance, or unavailability
- **Safety risk –** unsafe decisions affecting people, assets, or the environment
- **Compliance risk –** exposure to regulatory, contractual, or certification requirements
- **Supply-chain risk –** opaque models, undocumented dependencies, and inherited AI behavior
- **Reputational risk –** inability to explain or justify AI outcomes to stakeholders

What these risks have in common is that they are not always visible at the system or project level. Without systematic identification, classification, and governance, they remain scattered across departments, suppliers, and technologies.

Figure 1: Five Dimensions of AI Risk in the Technical Industry — AI risk typically spans operational, safety, compliance, supply-chain, and reputational dimensions, often invisible at the system or project level without systematic governance.

This is why trust in AI cannot be achieved through isolated technical measures alone. Only by maintaining an integrated view of AI usage, risk, and responsibility can organizations move from reactive control to proactive resilience; and ensure that AI remains governable as it scales in scope, autonomy, and impact.

From Insight to Action

The technical industry has never adopted new technology by assuming best-case behaviour. Safety, quality, and compliance have always been achieved through structure, evidence, and independent assurance. AI should be no exception.

As AI systems become more autonomous, more embedded, and more impactful, organizations must move beyond experimentation and fragmented controls. The path forward is not to wait for regulation to dictate action, nor to rely on technology teams alone, but to deliberately embed AI into existing management, safety, and governance structures; where risk is understood, responsibilities are clear, and trust is demonstrable.

Nemko Digital can help you to treat AI as industrial infrastructure. Map where it is used, understand what it influences, govern it across its lifecycle, and ensure that safety, supply-chain responsibility, and assurance are built in from the start. In that way we support organizations to be better prepared for regulatory and market expectations, and to scale AI with confidence, credibility, and resilience. Reach out to our experts to explore what this could mean in your situation.

About the author

dr. Pepijn van der Laan

Global Technical Director, AI Governance | Nemko Group
With two decades of experience at the intersection of AI, strategy, and compliance, Pep has led groundbreaking work in AI tooling, model risk governance, and GenAI deployment. He has advised multinational organizations on scaling trustworthy AI—from procurement chatbots to enterprise-wide model oversight frameworks.

Nemko Digital