Nemko Digital

Our New Year AI Trust Special

2025 in Retrospect, 2026 in Focus

2025 in Retrospect, 2026 in Focus

Bas Overtoom
Global Business Development Director

Pepijn van der Laan
Global Technical Director, AI Trust

As we are getting ready for Christmas and New Year, it feels like the right moment to pause, take a breath, and look back at what has been an extraordinary year in the evolution of AI. 2025 has moved at remarkable speed, reshaping industries, challenging assumptions, and redefining what responsible and trustworthy AI must look like. **In this New Year AI Trust Special, we will reflect on the past year through a timeline of key developments and shifts in global AI regulation, before we turn our eyes to 2026.** There we see four themes taking centre stage: 1) the rise of strategic and responsible procurement of AI solutions, 2) the need to scale AI governance with the right tooling and technology, 3) the growing urgency of bringing AI Agents under control, and 4) a continuous regulatory dynamic that will influence the landscape worldwide.

Looking back on 2025, what stands out most is how quickly the field has matured. **In the space of a single year, AI moved from exploration to large scale deployment.** Enterprise GenAI systems and AI Agents are now becoming firmly embedded into company workflows, customer interactions, and backend processes. This shift has brought a new level of urgency around AI trust, safety, and control. **It is not surprising that ISO 42001 adoption accelerated across industries. Many organizations now recognise that trustworthy AI cannot rely on goodwill or informal processes.** It requires disciplined management systems, clear documentation, and operational safeguards.

This deeper adoption also pushed the underlying infrastructure forward. **AI agent** communication frameworks matured significantly, and **MCP (Model Context Protocol)** has become the practical standard for enabling controlled interactions between different agents. Meanwhile, **multimodal and on-device AI capabilities** advanced faster than expected. The focus across the industry shifted away from model size and raw power towards specific **functionality, reliability, controllability, and energy efficiency.** Beyond digital environments, AI adoption in physical products accelerated, creating new expectations around lifecycle safety, software updates, and post-market monitoring. All of this unfolded alongside major regulatory developments. From the **EU AI Act** to digital omnibus initiatives and growing global alignment, 2025 showed that compliance is no longer a distant concept but a real operational requirement.

For Nemko Digital, it was also a dynamic and defining year. We had the honour of participating in **Dubai AI Week**, which opened the door to our first projects in that region. Our collaboration in **Korea with KSA (Korean Standards Association)** progressed from concept to execution as we shaped the **Nemko AI Trust Mark** and began bringing it to market. We co-hosted the **AI Trust in Electronics Summit with IBM in Oslo** and deepened our partnership with them. Lastly, we expanded our team, broadened our client portfolio across sectors, and moved into a new office in Amsterdam South. We are proud to see that we are increasing our impact in line with our mission to **Provide Trust in a Digital World**.

So, as we begin this New Year AI Trust Special, let us look back with clarity and forward with purpose. The past year has brought extraordinary change. The coming year will demand even more maturity, collaboration, and leadership. Today gives us the chance to explore where we stand, what we have learned, and where we are heading next.

1. Reflecting on 2025

AI moved fast in 2025, with new technologies, regulatory milestones, and industry shifts landing almost every month. It's impossible to capture everything that happened, but the events we selected here reflect the ones that stood out globally and had a meaningful impact on how AI is being built, governed, and used. Together, they give a clear picture of how the landscape evolved, and what organizations need to be aware of as they think about AI trust, compliance, and responsible adoption moving forward.

AI in 2025: Regulations Maturing and AI Becoming Autonomous

Key Technology Developments, Regulation Releases and Safety Aspects at a Glance

**See the Appendix for more detailed analysis.*

January 2025

Stargate Project announced (OpenAI, SoftBank, Oracle, MGX)

January 2025

First International AI Safety Report published

February 2025

First wave of EU AI Act obligations takes effect

February 2025

AI Action Summit, Paris

March 2025

OpenAI launches dedicated agent tools and SDKs

March 2025

WHO issues Guidance on AI in Health

May 2025

Anthropic launches the “Claude 4” Family (Opus & Sonnet)

July 2025

ISO/IEC 42006:2025 published

August 2025

OpenAI launches GPT-5

October 2025

OpenAI valued at ~$500B

November 2025

Draft CEN/CENELEC AI Quality Management System (QMS) standard (prEN 18286) released

November 2025

Google DeepMind releases Gemini 3 Pro

November 2025

Next version of the Model Context Protocol (MCP) specification

December 2025

Anthropic donates MCP to the new Agentic AI Foundation under the Linux Foundation

What this shows about the year in AI:

Looking at this timeline of events, it's clear that 2025 was a turning point for AI. The year felt like AI truly went global. From boardrooms to international summits, governance shifted from good intentions to hard requirements. With autonomous agents emerging, models advancing faster than ever, and massive investments reshaping the landscape, keeping AI trustworthy and safe became a real, urgent challenge for every organization.

1. Infrastructure and market concentration became critical themes

Projects like Stargate and OpenAI's $500 billion valuation show AI power consolidating among a few global players. At the same time, nation states are investing heavily in AI, with the United States leading (~$470 billion), China (~$119 billion), and smaller but significant commitments from Canada and the EU. These concentrated investments create both dependency risks and opportunities for structured assurance, vendor oversight, and multi-provider governance strategies.

2. Trust and compliance shifted from "good practice" to mandatory infrastructure

With the EU AI Act now enforceable and the development of CEN/CENELEC harmonized standards still in progress, organizations face growing obligations alongside emerging ISO standards and sector-specific guidance from WHO. This makes predictive, proactive AI governance essential to manage recurring, externally scrutinized compliance and risk requirements.

3. AI became a fully globalized strategic priority

Major summits, international reports, and cross-border investments demonstrate that AI has become a geopolitical and economic priority on par with climate, energy, and national security. Governance expectations are now shaped internationally, not only domestically.

4. Model capabilities advanced faster than governance cycles

Major releases (Gemini 3.0, Claude 4.5, GPT-5) arrived rapidly, each expanding capability boundaries. The pace highlights the importance of continuous evaluation, dynamic governance, and scenario-based risk management, not static compliance checklists.

5. Agentic AI moved from theory to early real deployment

2025 marked the transition from simple assistants to task-oriented, multi-step autonomous agents, with commercial launches, emerging protocols, and early foundations for interoperability. This shift requires rethinking risk — focusing not only on model outputs but on model-initiated actions and tool use.

2. State of Regulations in 2025

Looking across 2025, AI regulation matured into a complex, globally diverse landscape. Some regions pursued ambitious, enforceable frameworks, while others relied on voluntary guidance or sector-specific rules. Together, these approaches paint a nuanced picture of how different jurisdictions are shaping AI governance. Let's explore the regional dynamics in detail below.

Global Landscape

In 2025, the global regulatory environment for artificial intelligence continued to mature, driven by efforts to balance innovation, safety, human rights, and economic competitiveness. Instead of a single global standard, what emerged was a patchwork of regional and national frameworks, informed by shared principles but differentiated in implementation and emphasis.

Across regions, regulators grappled with common themes: risk-based classification of AI systems, transparency and accountability requirements, data governance, and procedural safety mechanisms. International bodies such as the **OECD** and the **Global Partnership on AI (GPAI)** continued to promote principled approaches to trustworthy AI aligned with human rights and democratic values, while enabling responsible innovation.

Several jurisdictions advanced major new laws and requirements in 2025. For instance, **South Korea's Basic AI Act**, formally *the Act on the Development of Artificial Intelligence and Establishment of Trust*, was promulgated in January 2025 and establishes a comprehensive legal and policy framework for AI across risk tiers and reinforcing transparency and accountability expectations. **In China**, regulators introduced a new AI content labeling regime, effective September 1, 2025, requiring providers to label AI-generated text, images, audio, video, and virtual scenes **to combat misinformation and improve traceability**, alongside additional national standards on generative AI security taking effect later in the year.

Different countries adopted varying strategies: some prioritized comprehensive, enforceable legislation (as seen in the EU), others took sectoral or risk-tiered regulatory approaches, and many lower-capacity jurisdictions leaned heavily on soft law, standards, and guidelines to avoid stifling domestic AI ecosystems. **A key trend in 2025 was the expansion of cross-jurisdictional dialogues and capacity-building initiatives**, seeking interoperability and mutual recognition of compliance regimes — but without convergence on a singular global standard.

This global mosaic set the backdrop against which regional powers competed and cooperated, influencing everything from trade policy and data flows to investment and geopolitical competitiveness.

Read more:
Global AI Regulations:
2025 Overview and Key Frameworks

United States

In the United States, 2025 was marked by an absence of a coherent federal AI regulatory framework. Despite mounting discourse around AI's societal impacts — from bias and misinformation to economic displacement — Congress failed to pass comprehensive AI legislation, largely due to sustained lobbying efforts by major technology firms.

A detailed analysis of the US landscape shows that **big tech influence** played a central role in this outcome. Lobbying expenditures and political action committees backed industry-friendly candidates and policies, championing a framework of "permissionless innovation" that avoids pre-emptive regulation and emphasizes market-led development. This approach prioritized industry flexibility and competitive advantage, particularly vis-à-vis China, over regulatory certainty or strict public-interest safeguards.

The current regulatory reality is highly decentralized and patchwork:

- At the federal level, policymakers leaned toward broad executive guidance and voluntary frameworks rather than binding legislative obligations.
- Several states introduced their own AI laws, but these efforts remained uneven and limited in scope, creating compliance complexity for companies operating across jurisdictions.
- Attempts by the U.S. House to impose extensive pre-emption of state AI regulation were watered down in the Senate, highlighting the political tension between federal primacy and state experimentation.

Consequently, American AI regulation in 2025 offered limited protections and guidance, leaving substantive governance to industry self-regulation, civil litigation, and consumer pressure — a landscape that critics argue exacerbates risks such as bias in automated decision-making and opaque data practices.

Read more:
How Big Tech Lobbying
Stopped US AI Regulation in 2025

European Union

In late 2025, the EU Commission proposed a one-year AI Act delay for high-risk systems through unveiling the **Digital Omnibus** package. It is a comprehensive initiative aimed at simplifying and updating the EU's digital rulebook, including the AI Act, GDPR, cybersecurity, and data governance frameworks. The package responds to concerns that the EU's ambitious digital legislation had become overly complex and administratively burdensome, particularly for smaller companies. Its objectives are twofold: to support innovation and competitiveness while maintaining the EU's strong focus on trust, safety, and human rights.

Key elements of the proposal include:

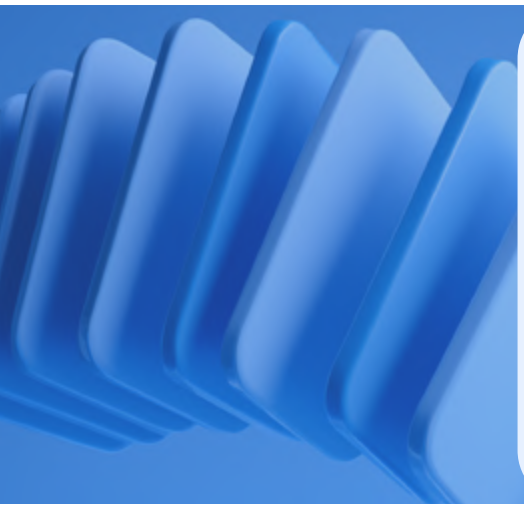
- **Delays to high-risk AI compliance deadlines by 1 year**, giving companies additional time to implement required safety, transparency, and accountability measures, and aligning implementation with forthcoming, yet delayed, technical standards.

- Flexibility for SMEs and mid-sized companies, including simplified documentation and compliance procedures, reducing the administrative load without compromising core safeguards.
- Unified reporting mechanisms for data and cybersecurity incidents, consolidating obligations under GDPR, NIS2, and the AI Act into a single, streamlined portal.
- Revisions to GDPR-related AI provisions, clarifying how personal data can be used for AI training and the conditions under which pseudonymised data may be processed.

The proposed package received mixed reactions. Industry groups generally welcomed it as a pragmatic effort to reduce red tape and support innovation. Civil society organizations, however, warned that revising GDPR and AI Act rules could weaken fundamental privacy protections and erode accountability in AI deployment. These tensions highlight the EU's ongoing challenge: reconciling values-based regulation with economic competitiveness in a fast-moving AI landscape.

The Digital Omnibus now enters the legislative negotiation phase with the European Parliament and Council, and discussions throughout 2026 will determine its final shape. Once adopted, it is expected to guide the EU toward a more streamlined, flexible, and pragmatic regulatory framework, positioning Europe as a global leader in trustworthy AI while adapting to practical realities on the ground.

Read more:
EU AI Act Delay Announced – Compliance Timeline Extended



Conclusion on the AI regulatory landscape

By the end of 2025, AI regulation had matured into a globally diverse terrain: the US with light, industry-shaped governance; the EU with ambitious yet contested regulatory refinement; and the broader global landscape populated by emerging frameworks seeking to balance innovation with rights protections. These divergent paths reflect broader geopolitical, economic, and societal priorities that will continue to shape AI policy evolution into 2026 and beyond.

3. The outlook

Four Trends That Will Shape AI Governance and Enterprise Readiness in 2026

As organizations prepare for the next wave of AI adoption, it is becoming clear that 2026 will not simply be a continuation of 2025. It will mark a shift toward more structured, more accountable and more scalable AI operations. GenAI, autonomous workflows and enterprise-wide deployment are moving from experimentation into core business processes, and with that shift comes a new set of responsibilities. Across industries, and particularly in domains where reliability, safety and quality are essential, four major trends will define how organizations navigate the coming year.


AI Trust:
Key Trends
for 2026
Fast-Moving AI Regulatory Landscape
Rapid expansion of global AI regulation (EU AI Act, product-related AI rules, country-specific laws).
Accelerating timelines increase demand for structured, repeatable compliance approaches.
Continuous monitoring and clear ownership are essential to manage regulatory dynamics.
Scaling AI with Tools & Technology
Shift from isolated pilots to enterprise-wide AI governance platforms.
Tooling required for risk management, documentation, monitoring, and compliance evidence.
Automation is critical to support audits, model inventories, and scalable AI deployment.
Autonomous AI Under Control
Growing autonomy requires stronger yet adaptable controls, guardrails, and oversight.
Key risks include unsupervised actions, cascading hallucinations, data leakage, and cyber exposure.
Clear definitions needed for permitted actions, human-in-the-loop, and continuous monitoring.
Responsible AI Procurement
AI procurement requires structured due diligence on risk, data usage, and compliance posture.
Evaluation should cover model origin, training data, security, robustness, and update processes.
Contractual clarity is essential to define roles, responsibilities, and liability across the value chain.

1. A Regulatory environment in motion

Around the world, AI regulation is evolving quickly, and 2026 will be a year where organizations must actively keep pace with shifting expectations. In Europe, **the EU AI Act continues to move toward implementation**, even as discussions progress around stretching some high-risk system obligations into 2027. This **proposed delay should not be misinterpreted as regulatory easing**. Instead, it reflects a desire to ensure that harmonised standards, conformity pathways and technical guidelines are strong enough for full enforcement. It creates a short window for companies to prepare, strengthen their documentation and governance, and build the trust by design maturity they will ultimately need.

At the same time, other European regulations such as the **Data Act** and the **updated Machinery Regulation** will add concrete obligations for organizations that integrate AI into products, production tools and connected systems. These rules, together with updates in product safety law, will significantly raise expectations around lifecycle monitoring, update management and transparency.

Importantly, these developments are not exclusive to Europe. **Korea's Basic AI Act has emerged as one of the most forward-looking national frameworks**, setting clear expectations on transparency, safety and accountability. Its structure has resonated globally because it blends innovation incentives with operational requirements, a combination increasingly relevant to multinational enterprises. **The Act is scheduled to come into force in January 2026**. In the United States, regulatory activity has been accelerating in a decentralized fashion through federal agency mandates, executive guidance, and state-level rules on fairness, safety, and data governance. However, in December 2025, the federal government signaled its intent to limit conflicting state AI laws through **Trump's executive order**, aiming to block states from regulating AI companies and strengthen U.S. leadership in AI through low-burden national policy. Layered on top of this is the growing influence of the **NIST AI Risk Management Framework**, which is shaping corporate governance practices even where formal AI law does not yet exist. In 2025, NIST also updated related foundational frameworks to support AI governance, releasing a draft Privacy Framework 1.1 with AI risk provisions and alignment to the Cybersecurity Framework, and launching SP 800-53 control overlays for securing AI systems. Together, these efforts make risk and security frameworks more interoperable and relevant for AI.

This creates a regulatory environment that is dynamic rather than fixed. Companies must watch closely for updates to harmonised standards, early interpretations by national regulators and sector-specific guidance that may emerge with little notice. They should anticipate growing expectations for live system monitoring, more structured update processes and clearer accountability across the supply chain. They will also need strong internal ownership over AI compliance, with clearly defined roles for risk management, monitoring and decision-making. In this climate, the organizations that succeed will be those that treat regulatory readiness not as a compliance task but as a strategic capability.

2. Scaling AI Governance with modern tooling

As AI adoption expands across the enterprise, organizations can no longer rely on manual tracking, fragmented registers, or isolated risk assessments. The movement from pilots to scaled deployment brings an unavoidable reality: **governance must modernize at the same pace as the technology itself**. In 2026, the most mature organizations will increasingly lean on dedicated AI governance platforms that support model inventories, risk evaluations, data lineage tracing, monitoring for drift and bias, and automated evidence generation for audits.

Modern AI governance tooling is becoming more sophisticated. Inventories are evolving to support agent catalogs, enabling organizations to track not just models but also autonomous agents and their interactions. Monitoring and evaluation capabilities are integrating more deeply, providing continuous oversight across performance, fairness, and security metrics. Platforms increasingly offer cross-platform and enterprise-wide compatibility, including the ability to manage Edge AI deployments, ensuring coverage for distributed or offline systems. Compliance workflows are also being streamlined, allowing automated capture of evidence for audits and regulatory reporting, reducing manual burden while enhancing transparency.

This reflects a broader industry understanding that **trustworthy AI cannot be achieved by human oversight alone**. With AI systems embedded in dozens – or in some cases hundreds – of applications, tooling becomes essential for achieving consistency. It allows organizations to see, for the first time, how all their AI systems behave over time, how they are updated, and where new risks may be emerging. It transforms governance from a collection of documents into an operational capability that runs continuously and at scale.

The shift is significant because regulators are increasingly expecting exactly this type of continuous control. Compliance will no longer be defined by annual reporting or static documentation, but by an organization's ability to monitor models in real time, intervene when issues appear, and demonstrate that controls are active rather than theoretical. By investing in modern governance tooling, organizations are not only preparing for compliance – they are enabling safe and confident expansion of AI across business units, markets, and technology stacks, while addressing the growing complexity of AI ecosystems, including agents, edge deployments, and multi-platform integration.

3. Bringing autonomous AI systems (AI Agents) under control

The third trend shaping 2026 is the rapid expansion of autonomy across enterprise workflows. Isolated **AI agents** are evolving into broader ecosystems of autonomous and semi-autonomous systems that can initiate actions, interact with tools, and influence outcomes without direct human prompting. In industrial settings, this includes AI managing operational workflows, robotics adapting to real-time conditions, and equipment making decisions previously requiring manual oversight. These systems promise higher efficiency, improved consistency, earlier risk detection, and enhanced production quality.

However, **autonomous AI introduces new categories of risk**. Systems can move beyond their intended scope, chain multiple steps unpredictably, amplify errors faster than humans can detect, and interact with each other in ways that create emergent behavior. When embedded in physical systems—such as industrial machinery, vehicles, or critical infrastructure—these risks directly impact functional safety, potentially causing harm if AI decisions lead to unsafe actions. Autonomous systems may also expose data unintentionally, trigger undesired operations, or obscure the decision path behind errors, complicating accountability and incident response. **Managing these risks is becoming a core enterprise competency**.

To address this, organizations are adopting structured governance approaches. **The Model Context Protocol (MCP) is emerging as a foundational standard**, defining what tools AI systems may access, operational boundaries, and requirements for auditable actions. MCP constrains both technical capabilities and intent, ensuring alignment with enterprise security, authorization, privacy, and safety expectations.

Complementing MCP, enterprise autonomy frameworks translate organizational risk appetite into operational rules. They define which tasks can be autonomous, when human oversight is required, and the approval processes for production deployment. These frameworks clarify where humans stay in the loop, remain on the loop, or step aside, balancing independence with supervision.

Monitoring and observability are now essential. Continuous insight into AI behavior, tool usage, anomalies, and deviations is critical, alongside real-time intervention capabilities to pause, rollback, or stop systems. These safeguards rely on robust identity and access management, treating AI systems like privileged users with distinct identities and permissions within a zero-trust architecture.

As enterprises deploy multiple interconnected autonomous systems, **cross-system control becomes the next frontier**. Governing collaboration, information exchange, and conflict resolution is crucial to prevent unintended emergent outcomes. Effective oversight requires orchestration, supervision models, and holistic observability across the ecosystem.

In 2026, autonomous AI will drive efficiency while introducing operational risk. Organizations that succeed will combine ambition with discipline, leveraging standards, governance frameworks, and monitoring to scale autonomy safely, transforming it from a source of uncertainty into a competitive advantage.

4. Responsible AI procurement as a strategic capability

The final trend shaping 2026 is perhaps the most underappreciated. AI is now entering organizations primarily through **procurement** rather than internal development. This shift changes how companies must think about AI risk. It means that trust, safety and compliance must be established before a system enters the organization, not retrofitted afterwards.

In practice, **this requires a deeper level of scrutiny than many organizations are used to**. Companies must understand where a model originates, what data it was trained on, how it performs across different contexts, and how it will be updated throughout its lifecycle. They must evaluate supplier maturity, including the quality of documentation, the transparency of model cards and datasheets, the strength of security measures, and the rigor of testing processes. **Suppliers** should also be assessed on their AI risk ratings, which indicate their overall posture across governance, safety, and reliability. **Organizations must establish clear roles and responsibilities**, define accountability for incidents, ensure transparency obligations are met, and negotiate contract clauses that reflect the operational realities of AI systems. This becomes even more important when integrators or consultancies develop AI solutions on a client's behalf, as these often arrive with limited visibility into long-term support, monitoring commitments, or update pathways. Under the EU AI Act, organizations bear supply chain responsibility for high-risk AI systems, which increases complexity and operational costs for providers, while tools like **Nemko Digital's AI Trust Mark** offer an efficient way to demonstrate trustworthiness across governance, safety, and operational standards.

Responsible procurement is therefore becoming one of the most powerful risk management levers in the enterprise. It determines which systems enter the organization, under what rules, with which guarantees and with what rights to inspect, monitor and intervene. It enables companies to set transparent expectations around bias testing, update management and lifecycle monitoring before a system is even accepted. And it ensures that third-party systems, which often carry the highest level of uncertainty, are brought under the same governance umbrella as internally developed systems.

As AI becomes integral to products, production processes and digital services, procurement will no longer be viewed as a transactional function. It will become a strategic cornerstone of corporate governance, central to the organization's ability to deploy AI safely, responsibly and at scale.

4. Conclusion

As we close this New Year Special, one message becomes unmistakably clear. **We are entering a period where AI is no longer a frontier technology but a foundational one**. The developments of 2025 have shown that organizations across the world are moving beyond experimentation and are now relying on AI in ways that shape core operations, societal interactions, and entire market segments. **With this shift comes a shared responsibility**: to ensure that the systems we build are reliable, fair, transparent, and safe. The acceleration of ISO 42001, our own **Nemko AI Trust Mark** and other related standards and regulations signal a growing recognition that good intentions are not enough. Trust must be verifiable, operationalised, and embedded throughout the lifecycle of AI.

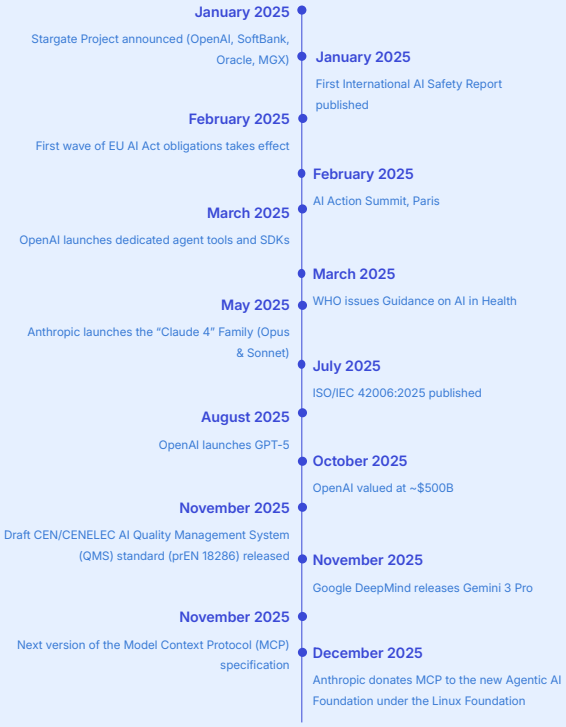
The coming year will bring even more complexity, but also unprecedented opportunity. **Procurement** processes will increasingly become crucial to meeting quality and compliance expectations. **Governance and monitoring** will have to scale alongside technology. **AI Agents** will require new forms of control, oversight, and coordination. **Regulatory frameworks** will continue to evolve and converge, reflecting a world that is trying to keep pace with innovation while protecting societal interests. These changes demand leadership, collaboration, and a willingness to make deliberate choices about how AI is built and deployed.

For Nemko Digital, the path forward is clear. **We enter 2026 with a stronger team, a broader portfolio, and a mission that matters more than ever**. The work we do is no longer niche or optional. It is central to how organizations earn trust, reduce risk, and unlock sustainable value from AI. Together with our clients, partners, and global network, we will continue shaping a future where innovation and responsibility reinforce one another.

Thank you for your engagement throughout the year, and for your interest and commitment to building a trustworthy digital world. We look forward to continuing this journey with you in 2026. Merry Christmas and Happy New Year!

Nemko Digital

Appendix



January 2025 — Stargate Project announced (OpenAI, SoftBank, Oracle, MGX)

What happened:

The Stargate Project, a major private AI infrastructure initiative, was launched as a joint venture between OpenAI, SoftBank, Oracle, and MGX. It involves multi-billion-dollar investments in next-generation data centers, GPU/TPU clusters, and specialized AI compute for frontier-scale models, emphasizing supply-chain security and long-term capacity planning. It showed how frontier AI now depends on centralized, tightly governed compute, making strong assurance and third-party risk oversight increasingly essential.

January 2025 — First International AI Safety Report published

What happened:

The first International AI Safety Report mapped systemic AI risks, misuse pathways, and governance gaps across cyber, biosecurity, autonomy, misinformation, and economic domains. Led by Turing Award winner Yoshua Bengio, authored by 100+ experts, and backed by 30 countries alongside the UN, OECD, and EU, it became the largest global collaboration on AI safety to date and was widely cited ahead of major 2025 policy decisions. It reinforced AI safety as a global priority and highlighted the need for strong risk assessments, impact evaluations, and continuous monitoring.

February 2025 — First wave of EU AI Act obligations takes effect

What happened:

The earliest provisions of the EU AI Act took effect, including bans on prohibited AI practices, transparency requirements, and initial risk controls for high-impact systems. Companies operating in or serving the EU had to adjust documentation, system design, and governance processes to comply. This marked the shift from AI governance as a concept to legally enforced reality, showing businesses that proactive compliance is essential.

February 2025 — AI Action Summit, Paris

What happened:

The Summit brought together over 100 nations, major tech firms, research leaders, and civil society to coordinate global AI strategy. It produced the Statement on Inclusive and Sustainable AI, launched initiatives on public-interest AI, future of work, and equitable access to compute, and pledged capacity-building in developing countries. The event established that AI trust, compliance, and responsible deployment are becoming transnational expectations, signaling that companies must align internal policies not only with local laws but also with emerging international norms and trust standards.

March 2025 — OpenAI launches dedicated agent tools and SDKs

What happened:

OpenAI released dedicated tools and SDKs enabling developers and enterprises to build autonomous AI agents capable of planning, reasoning, and executing multi-step workflows across software systems. The release accelerated adoption of agentic AI, provided standardized frameworks for integrating AI with third-party services, and offered monitoring and control features to manage autonomous behaviors. This signaled that agentic AI was becoming widely accessible and operationalized, highlighting new risks and compliance needs around autonomy, auditability, delegation boundaries, and safe-action constraints for companies deploying such agents.

March 2025 — WHO issues Guidance on AI in Health

What happened:

The WHO published guidance on governance, ethics, and safe

use of large multimodal AI models in healthcare and public health, covering clinical validation, bias mitigation, human oversight, data governance, and safety testing for high-stakes applications. It highlights growing global alignment on clinical-grade AI compliance, requiring vendors, hospitals, and insurers to meet rising standards for evidence, documentation, and safe deployment.

May 2025 — Anthropic launches the "Claude 4" Family (Opus & Sonnet)

What happened:

Anthropic released Claude Opus 4 and Sonnet 4, featuring improved reasoning, code generation, and safety evaluations guided by Constitutional AI frameworks. The launch included detailed system cards and safety disclosures, setting transparency benchmarks. By November 2025, the Claude family had advanced to Claude 4.5, with enhancements in agentic capabilities, coding performance, and efficiency compared to 4.0, reinforcing expectations that AI developers provide robust safety information and operational improvements.

July 2025 — ISO/IEC 42006:2025 published

What happened:

ISO 42006:2025 established requirements for entities auditing and certifying AI management systems, complementing ISO/IEC 42001. It created a formal global ecosystem for accredited AI governance audits. The standard strengthens AI assurance infrastructure, enabling organizations to demonstrate trustworthiness to regulators, partners, and clients, and signaling a future where AI governance maturity is externally verifiable, similar to ISO 27001.

August 2025 — OpenAI launches GPT-5

What happened:

OpenAI rolled out GPT-5 globally, with improvements in multimodal reasoning, agentic capabilities, and controllability. It quickly integrated into productivity platforms, coding assistants, and enterprise applications. The release underscores the fast pace of high-capability models, requiring continuous governance, vendor reassessment, and careful validation by organizations.

October 2025 — OpenAI valued at ~$500B

What happened:

OpenAI's valuation reached approximately $500 billion, becoming the world's most valuable private tech company. Growth was driven by enterprise adoption, ecosystem integrations, and its role in defining the frontier AI platform. The milestone highlights concentration and dependency risks, requiring enterprises to manage supply-chain exposure, vendor transparency, and governance alignment.

November 2025 — Draft CEN/CENELEC AI Quality Management System (QMS) standard (prEN 18286) released

What happened:

CEN and CENELEC released the draft standard prEN 18286: Artificial Intelligence – Quality Management System for EU AI Act Regulatory Purposes for public consultation, which is open for comments until 22 January 2026. The draft outlines QMS requirements, including organizational processes, risk management, documentation, and lifecycle controls to support compliance with the EU AI Act. The draft signals growing harmonization of European AI governance and provides guidance for organizations to implement structured quality and compliance processes aligned with upcoming harmonized standards under the EU AI Act.

November 2025 — Google DeepMind releases Gemini 3 Pro

What happened:

Google DeepMind launched Gemini 3 Pro, the latest frontier AI model with state-of-the-art reasoning, multimodal understanding, and long-horizon planning, nearly doubling its predecessor's score on Humanity's Last Exam (37.5 % vs ~18.8 %) and outperforming Gemini 2.5 Pro across major benchmarks. It powers advanced workflows and agent integrations across Google products and developer platforms, highlighting the rapid pace of model advancement and the need for organizations to update risk assessments, governance frameworks, and compliance strategies for cutting-edge AI.

November 2025 — Next version of the Model Context Protocol (MCP) specification

What happened:

The Model Context Protocol (MCP), an open standard for integrating large language models with external data, tools, and applications, released its next version on 25 November 2025 following a mid-November release candidate. It introduced enhanced authorization, structured operations, governance structures, and community-driven extensions. MCP continues to gain adoption as a backbone for interoperable AI systems, standardizing how agents connect with workflows and data across platforms. The update strengthens MCP's role in secure, standardized AI connectivity and highlights the need for organizations to embed such interoperability protocols into governance and risk frameworks as agentic AI scales.

December 2025 — Anthropic donates MCP to the new Agentic AI Foundation under the Linux Foundation

What happened:

The widely adopted Model Context Protocol (MCP) was donated to the newly formed Agentic AI Foundation under the Linux Foundation. The Foundation guides standards, interoperability, and safe deployment for agentic systems, supports community-driven development, and helps define open governance for AI protocols. This shift toward ecosystem-level interoperability requires organizations to manage compliance, interface security, and clear responsibility boundaries as agentic AI increasingly operates across platforms.


Nemko
Digital