Responsible AI

WHITEPAPER

Embed responsible AI to drive value, mitigate risk, and accelerate innovation

ust.com

UST

Table of contents

Executive summary

Three years after the launch of ChatGPT, we're experiencing an explosion of interest in AI across industry and society, and we're reaching a pivotal point in the hype curve. Many AI initiatives are failing to progress beyond the proof-of-concept (PoC) stage into live service, while others are succeeding and generating significant benefits.

Enterprises are realizing that AI must be carefully controlled and risk-managed, as uncertainty is a major factor in project failures. The common perception is that being responsible with AI slows things down, when actually the converse can be true. Central to this is the effective management of risks associated with AI. When enterprises implement such processes within a responsible AI framework, they demystify what can be seen as opaque and complex, simplifying the process while ensuring their AI is fair, secure, cost-effective, and error-free.

WHAT IS RESPONSIBLE AI?

There is significant interest in developing and using AI across industries for good reason. Used effectively, AI can improve the quality of services and yield significant efficiencies. However, the output generated by AI can be invalid or incorrect – checks are necessary to mitigate the associated risks. AI can also appear prejudiced, lead to security and legal challenges, and pose additional risks. The processes to manage these risks have collectively become known as Responsible AI.

WHY DO WE NEED RESPONSIBLE AI?

AI is built on data that records societal information. We live in a society that, unfortunately, is inherently biased, and the data collected from and by society often reflects these biases. AI can reflect and magnify these biases; therefore, it needs to be managed to ensure it is fair.

When an LLM model generates content, it does so by predicting sequences of events. These predictions can be wrong. To put it simply, the LLM predicts the incorrect word or line of code and continues to do so through a sequence or chain of events. When this happens, the content created can appear highly relevant (in the sense that it is well written or logically structured with relevant code) despite being completely inaccurate. This has become known as AI hallucination.

Agentic AI has increased the risk posed by AI hallucination. As Google DeepMind founder Demis Hassabis discussed in March 2025, autonomous agents acting on content generated by AI can lead to significant errors unless effective guardrails are in place.

In addition to the technical issues that can arise when using AI, enterprises must

also address legal compliance concerns. There are practical steps businesses and other organizations can take to manage the risks of AI error and unintended consequences. It is entirely possible to innovate with AI and benefit from these technologies in responsible, legal, and secure ways.

AI regulation right now

Globally, governments have responded to the risks associated with the misuse of AI or mistakes caused by AI AI by developing guidelines for its responsible use and, in some cases, enacting AI regulations into law The current status of these for several jurisdictions is below.

Jurisdiction	Position on AI regulation (Q2 2025/26)
EU	EU AI Act came into force August 2024 and will be fully applicable in August 2026. 'The first-ever comprehensive legal framework on AI worldwide.'
UK	No enacted approach to AI regulation. Principles-based approach outlined in 2023 UK Government White Paper.
US	Biden's AI Executive Order was rescinded by the new administration in January 2025. No single comprehensive law – a combination of federal initiatives and state-level legislation.
Canada	Artificial Intelligence and Data Act (AIDA), drafted in 2022, terminated in January 2025 due to the proroguing of parliament.
Singapore	No specific AI laws or regulations in place. Toolkits and guides to the deployment of Responsible AI including the 2023 AI Verify Toolkit.
China	Interim Measures for Generative AI (2023) are in place. From September 2025 all AI-generated content will need to be labelled as such.
Global	UN established a multi-stakeholder High-Level-Advisory Board for AI in October 2023. OECD AI Principles (updated 2024). ISO/IEC 42001:2023 AI Management Systems released in 2023.

Given the myriad different AI guidelines and regulations, and the variation in their relevant status (enacted, redacted, in development or unlikely to be developed), this can be a confusing area, particularly for non-experts.

This can also be challenging for enterprises, which are often large companies that work across multiple jurisdictions. However, in the industry today, companies and other organizations that have established or are establishing AI risk management processes are typically doing so using a small number of guiding frameworks, including the EU AI Act, the NIST AI Risk Management framework, and ISO/IEC 42001:2023, the international standard for AI Management Systems.

The guidelines are becoming clearer, but enterprises are still struggling to productionize and gain value from AI

Although the regulations and guidelines governing AI are becoming clearer, their impact is not yet being felt across the industry. In June 2025, Gartner predicted that more than 40% of Agentic AI will be cancelled by the end of 2027. Research published by MIT in August 2025, which found that 95% of investments in GenAI have so far failed to yield measurable value, prompted considerable debate.

These are clear signs that we are entering the trough of disillusionment phase of the Agentic AI hype cycle.

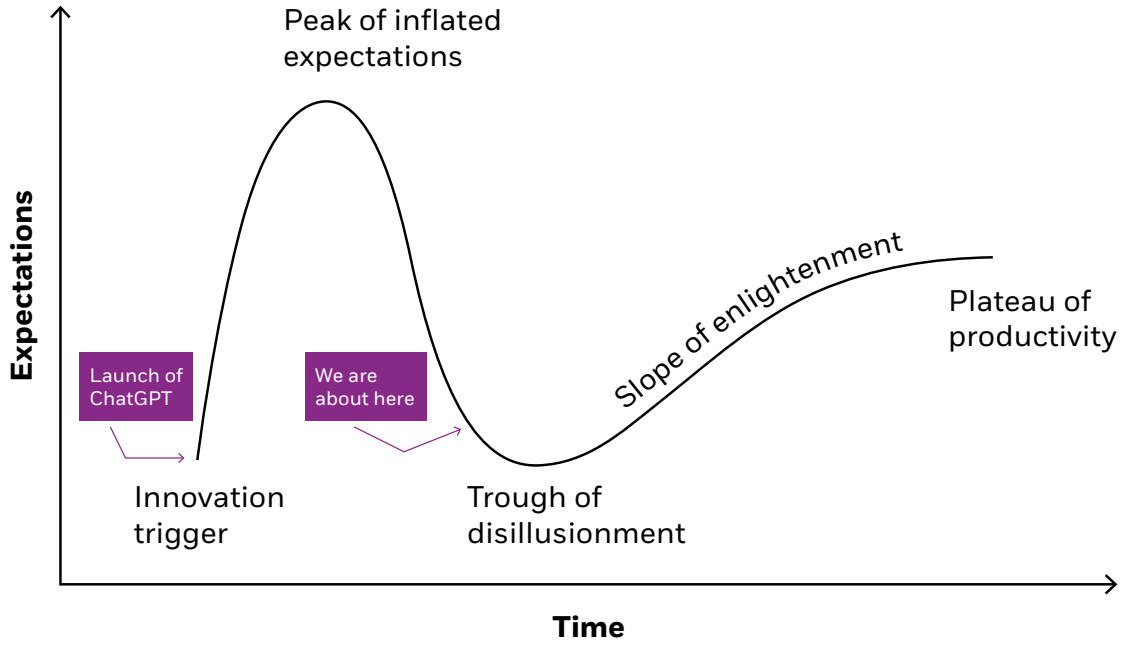


Figure 1: Agentic AI hype cycle

While realizing Return on Investment (ROI) is commonly cited, MIT reported the reasons behind this AI failure fall broadly into three groups:

- Lack of financial management leading to spiralling costs.
- Focus on gaining business value, not built into the plan for AI development and utilization.
- Lack of AI governance leading to risks associated with AI being poorly understood and thus unmanageable.

Despite these challenges, value is being achieved through the use of AI. The MIT research also makes it clear that enterprises that take a strategic approach to embedding and developing AI, with a clear focus on measuring ROI from the outset, along with controlling AI risk, are most likely to succeed with AI.

Attitudes toward trust in AI

In 2024, the UK government published a report, "Public Attitudes Towards AI Assurance," which researched and discussed the perceptions of AI assurance across UK society in general. This found that trust in AI systems was more likely to be based on the familiarity and trust individuals placed in a brand or service, rather than the AI itself. The majority of people have no idea how AI works; they often don't even know they are using it. By placing trust in the brand or service, they are entrusting their expectations that the AI will be accurate and behave fairly towards the service provider.

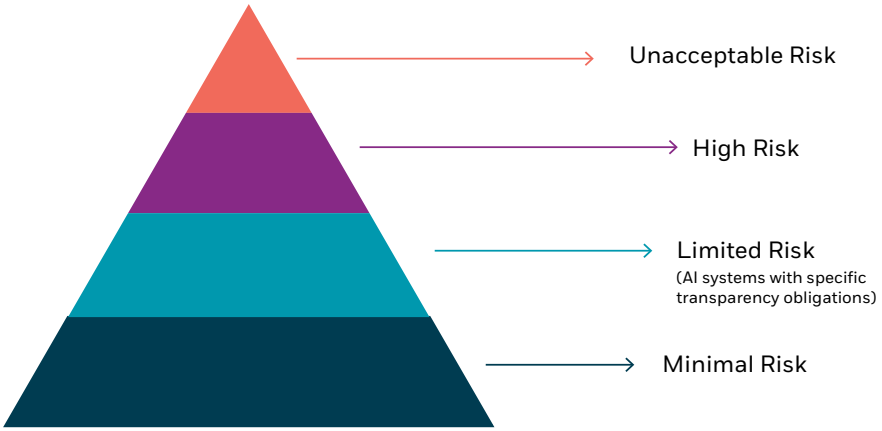
This places an implicit responsibility on enterprises that provision AI services to ensure these services are controlled. And it can go beyond reputation. Companies and organizations providing healthcare and similar services, for example, need to safeguard any AI they use within the patient care pathway to ensure it is unbiased, secure, and safe. Without such controls, adverse events can happen. For example, a Harvard Medical School article discusses the use of AI across several US healthcare systems that exhibited bias by prioritizing healthier white patients over sicker black patients due to the AI being trained on cost data and not care needs. While this is, of course, ethically wrong, healthcare companies that use these AI systems also open themselves up to litigation by patients whose health and welfare are adversely impacted due to the AI bias

Different scales of AI risk

It's important to recognize that many AI services in use within enterprises today are low risk. For example, when Amazon recommends me a book I'm not interested in, it doesn't really matter to me, I probably won't stop buying things on the platform. But it does matter to me if my child's exam grades are unfairly standardized by AI. These two examples illustrate how the risk associated with enterprises and public sector organizations using AI can vary.

Different AI services carry varying levels of severity if they malfunction. AI regulations and responsible AI frameworks recognize this; the EU AI Act is a notable example.

The regulatory framework defines 4 levels of risk for AI systems:



Source: EU AI Act AI Regulatory Framework: Click here

Such approaches facilitate a pragmatic approach to managing the risks associated with AI. Level 1 AI risk does not need any form of risk management, and mitigation for the given risk becomes more rigorous as the scale of adverse incidents caused by unfair or unsafe treatment increases. Whether an enterprise operates within a regulated jurisdiction, it can use such risk-based approaches to manage its AI.

The AI Council: Recording all AI in use and understanding its risk

Risks associated with AI can be categorized, as shown and detailed below .

Jurisdiction	Position on AI regulation (Q2 2025/26)
Technical risk	Security vulnerabilities, lack of model robustness, algorithmic bias, hallucination, software/model bugs.
Operational risk	Failures in deployment, integration and performance issues.
Ethical risk	Bias, discrimination, privacy violations, lack of transparency.
Legal and regulatory requirements risk	Non-compliance with laws and regulations. Intellectual property issues.
Social risk	Reduction in human resource requirements due to automation, scope for AI misuse, impact on society inequalities
Environmental risk	What is the energy consumption of the AI, and how does it impact the organization's ESG targets?
Geopolitical risk	What are the geopolitical risks associated with building and/or implementing this AI service.

Approaches to risk mitigation differ depending on the nature of the risk. Enterprises typically form an internal AI Council comprised of subject matter experts (SMEs) from across their legal, procurement, HR, AI, and wider business teams, who collectively have the expertise to understand and mitigate risks.

Taking this cross-functional approach has the added benefit of spreading ownership and responsibility for AI delivery across the business beyond data, ML, and IT teams, thus increasing the diffusion of AI across the enterprise.

AI register and risk logs

The AI Council is responsible for ensuring all AI in use across the business is recorded in its AI Register, with an associated risk log for all AI categorized as risk level 2 and above.

An example risk log for an AI Agent used for business intelligence reporting is shown below. The risk log also details the member of the AI Register responsible for mitigating the given risk, who will then take this forward in their respective business area.

AI Risk Types	Description	Likelihood	Impact	Mitigation
Technical	Algorithmic bias.	2	4	Ensure the LLM used in the Agent is suitably guard-railed. Include a human-in-the-loop check of all board reports, reports published externall and reports used for key business decisions.
Operational	Failure of integration into critical business systems caused by model drift.	2	2	Ensure MLOps/ LLMOps processes in place to monitor models for drift and decreasing accuracy.
Legal and regulatory	IP issues caused by LLM reproducing copyrighted code.	1	4	Use an LLM with assurances that no copyrighted code has been used to train the model.
Environmental	High use of energy by AI system.	1	3	Measure energy use of the AI system early in development. Assess whether this is appropriate given ESG targets. Report on energy use of the AI system when live.

This systematic approach to risk managing AI can be applied to any AI service and can be integrated into business processes. AI risk logs can be automatically generated for all standard forms of AI in use, simplifying and, to a large degree, demystifying AI assurance, making it less of a specialized field reliant on several key internal experts.

Systematic mitigation of AI risk - Embedding controls into AI production processes and live AI services

Mitigation steps for the majority of AI risks can be built into the systematic processes used to productionize and run the given AI service.

As enterprises increasingly mature their AI production processes (MLOps and LLMOps), AI controls such as data drift, evaluation, and bias checks, along with monitoring the energy use of an AI system, are embedded into these processes, with automated alerts triggering if the system trends outside stated limits and so carries risk. Gartner highlights the importance of AI observability in this context.

Red-teaming methods are also increasingly being used as a means of testing and assuring the rigor of an AI system before rogue agents do the same.

Along with compliance checks in business areas such as legal, HR, and procurement at the time of AI service development or procurement, these controls form a holistic approach that enterprises need to take to ensure they benefit from AI while maintaining control and oversight.

Measuring the benefits of responsible AI

Implementing responsible AI should enable enterprises to substantially minimize adverse incidents associated with AI. It's essential to maintain a record of adverse incidents, and this OKR should be reported and monitored as a proportion of deployed AI systems.

Succeeding with responsible AI: Case studies

MERCEDES BENZ

The automotive industry is undergoing significant disruption with the increased use of automation and AI. A global leader in this area, Mercedes-Benz, has implemented a four-principled approach: 1. Responsible use 2. Explainability 3. Protection of privacy 4. Safety and reliability across the entire manufacturing, sales, and support estate.

NATIONAL HEALTH SERVICE (NHS) – AI IMAGING RADIOLOGY PILOTS

It's widely recognized that AI has the potential to transform healthcare economies, which are under pressure from aging populations with an increasing number of morbidities. The shift from analogue to digital, much of which will be powered by AI, is detailed in the recently published NHS 10-year plan.

To realize these ambitions for AI, NHS leaders have acknowledged and are building in assurances that the AI in use is effective, clinically assured, and responsible. The use of AI to support diagnostic imaging in hospitals is currently undergoing trials, with complete transparency regarding how patient data is being used and how the AI is controlled.

Conclusion

Responsible AI is achievable, measurable, and can be automated to a significant degree, saving the time of enterprise SMEs and lowering the barriers to entry and achievement. Furthermore, research has shown that enterprises that implement Responsible AI across their business experience an increased rate of success with their AI services, while decreasing the likelihood of adverse AI incidents.

Over time, AI assurance processes will become as embedded and widely used as data governance controls within enterprises. The primary driver for data governance was the General Data Protection Regulation (GDPR), a 2018 EU regulation that, at the time, applied to the UK. While global AI regulations vary, the need for enterprises to ensure their AI services are trusted and secure remains.

Enterprises that adopt responsible AI today using clear and measurable approaches can gain a competitive advantage, as responsible AI is essential for realizing business value from AI.

Discover how your enterprise can embed responsible AI to drive value, mitigate risk, and accelerate innovation with confidence.

Connect with an expert →

Heather Dawe

Chief Data Scientist and Head of Responsible AI, UST

Heather Dawe is a Data and AI Leader with 25 years of experience working across the industry, innovating with data and how it can be used to improve outcomes, quality and efficiency. As Chief Data Scientist and Head of Responsible AI at UST, Heather works with global Enterprises in the financial services, retail, manufacturing and public sector, advising on Data and AI strategy and building associated implementation programmes.

Heather has appeared on the BBC, Sky News, and numerous national and international news publications including The Guardian, Financial Times and Economic Times. With her UST colleague Dr. Adnan Masood, she is co-author of the 2023 book Responsible AI in the Enterprise, a guide to how to succeed with AI in business safely, fairly and ethically. She spoke about AI ethics at the World Economic Forum in Davos in January 2025.

Disclaimer:

This whitepaper compiles information from publicly available sources and vendor materials as of the publication date, together with the author's analysis. It is provided "as is" for general information—not legal, financial, or professional advice—and no warranty is made as to completeness, accuracy, or timeliness; figures and product details may change, and references to third-party tools do not imply endorsement (all trademarks belong to their owners).

Together, we build for boundless impact

Since 1999, UST has worked side by side with the world's best companies to make a powerful impact through transformation. Powered by technology, inspired by people, and led by our purpose, we partner with our clients from design to operation. Our digital solutions, proprietary platforms, engineering, R&D, products, and innovation ecosystem turn core challenges into impactful, disruptive solutions. With deep industry knowledge and a future-ready mindset, we infuse expertise, innovation, and agility into our clients' organizations—delivering measurable value and positive lasting change for them, their customers, and communities around the world. Together, with 30,000+ employees in 30+ countries, we build for boundless impact—touching billions of lives in the process.

ust.com

Version 0103-20251124

U ▪
S T