# The rise of AI trust layers in addressing generative AI concerns



In the rapidly evolving landscape of artificial intelligence (AI), organisations are increasingly captivated by the potential of generative AI (GenAI). However, a significant disparity remains between the enthusiasm surrounding these technologies and their actual implementation. According to recent studies, only a small fraction of GenAI projects have transitioned into production, primarily due to organisations' concerns regarding the reliability of large language models (LLMs) and their propensity to "hallucinate" or produce inconsistent outputs.

Generative models, including LLMs, are distinguished by their ability to process vast quantities of unstructured data—from text documents to multimedia sources—and to respond to queries based on the insights derived from this data. This capability has been particularly attractive for applications such as chatbots, digital co-pilots, and numerous semi-autonomous agents capable of performing language-based tasks. Yet, as organisations embrace this technology, they remain cautious; LLM users often have limited control over the outputs generated, which can lead to answers that stray perilously from factual accuracy.

The challenge of managing the inherent unpredictability of LLMs, referred to as "hallucination" by some experts—including the National Institute of Standards and Technology (NIST)—has prompted businesses to explore the implementation of AI trust layers. This approach serves as a safeguard, enhancing the reliability of LLM outputs and ensuring a more consistent user experience. Notable industry players, such as Salesforce, have adopted various strategies within this framework. Their Einstein AI models employ secure data retrieval, dynamic grounding, data masking, and toxicity detection among other measures, thus minimising the likelihood of unsatisfactory customer interactions.

In light of the growing demand for a universal AI trust framework, Galileo, a company co-founded by Yash Sheth in 2021, is making headway in developing a trust layer that can function across multiple GenAI platforms and models. Before founding Galileo, Sheth gained substantial experience at Google, where his work on LLMs for speech recognition provided him with critical insights regarding the practicality of these technologies in enterprise settings. "We saw that LLMs are going to unlock 80% of the world’s information, which is unstructured data," Sheth remarked in an interview with BigDATAwire during the recent re:Invent event. However, he acknowledged the inherent difficulties in adapting these models for varied applications because of their non-deterministic nature.

Galileo aims to address this complexity by creating a robust AI trust layer that focuses on monitoring and managing LLM behaviour. The company's strategy involves employing a foundational model specifically designed to assess the performance of LLMs and developing a unique set of metrics to monitor their output. When these metrics indicate undesirable behaviour, preventative measures—referred to as guardrails—are enacted to avert potential issues. "These metrics that are computed each time in production and in real-time, in low latency, block the hallucinations, block bad outcomes from happening," Sheth explained.

The suite of tools offered by Galileo includes components such as Evaluate, which allows for experimentation within a client’s GenAI architecture; Observe, which ensures secure monitoring of LLM behaviour; and Protect, designed to prevent harmful requests and mitigate data leakage. Collectively, these tools facilitate trust in GenAI applications akin to confidence in traditionally developed software.

In less than two years since its inception, Galileo has started to carve out a niche in the market, securing clients from the Fortune 100, including notable companies like Comcast, Twilio, and ServiceNow. A partnership established with Hewlett Packard Enterprise (HPE) in July, along with a successful Series B funding round that raised $45 million in October, propels the company’s total venture funding to $68.1 million.

Looking ahead, as enterprises approach 2025, the need for AI trust layers is becoming increasingly critical. With many businesses eager to roll out their GenAI initiatives, operational directors remain hesitant until they can ensure the technology is reliable. Sheth is confident that Galileo’s strategy to mitigate the adverse effects of unpredictable AI systems is essential for fostering that confidence. He noted the transformative potential of GenAI and the importance of a well-structured trust layer in avoiding regression to outdated software practices. "There are amazing use cases that I’ve never seen possible with traditional AI," he said, highlighting the unique opportunities GenAI could unlock if those reliability concerns can be adequately addressed.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://bloggingwizard.com/generative-ai-statistics/> - Provides statistics on the global market size of generative AI, adoption rates, and the challenges of implementing generative AI, including the issue of 'hallucination' in large language models.
2. <https://explodingtopics.com/blog/generative-ai-stats> - Offers insights into the adoption rates of generative AI across various industries, the impact of generative AI on jobs, and the expectations of business leaders regarding its use.
3. <https://www.salesforce.com/news/stories/generative-ai-statistics/?bc=DB> - Discusses the adoption of generative AI by different populations, the frequency of use, and the trust levels in generative AI among users, which aligns with the need for reliable AI trust layers.
4. <https://bloggingwizard.com/generative-ai-statistics/> - Details the use of generative AI in various applications such as image creation, text creation, and customer service, highlighting the need for managing the unpredictability of LLMs.
5. <https://explodingtopics.com/blog/generative-ai-stats> - Mentions the expectation of business leaders to use generative AI for low-value tasks and customer service, which requires reliable AI trust layers to ensure satisfactory interactions.
6. <https://bloggingwizard.com/generative-ai-statistics/> - Highlights the urgency felt by executives to adopt generative AI despite lacking the necessary skills, underscoring the need for a well-structured trust layer.
7. <https://www.salesforce.com/news/stories/generative-ai-statistics/?bc=DB> - Discusses the widespread use of generative AI among different demographics and the importance of trust in these technologies, which Galileo aims to address.
8. <https://explodingtopics.com/blog/generative-ai-stats> - Provides statistics on the potential job impact of generative AI and the need for reliable AI systems to manage this transition effectively.
9. <https://bloggingwizard.com/generative-ai-statistics/> - Explains the adoption rates of generative AI in marketing and advertising, which is a sector that would benefit from robust AI trust layers.
10. <https://explodingtopics.com/blog/generative-ai-stats> - Mentions the piloting of generative AI programs by a significant percentage of organizations, indicating a growing need for reliable AI trust frameworks.
11. <https://www.salesforce.com/news/stories/generative-ai-statistics/?bc=DB> - Highlights the frequent use and trust in generative AI among younger generations, emphasizing the importance of ensuring the reliability of these systems.
12. <https://www.bigdatawire.com/2025/01/07/developing-a-trust-layer-for-ai-systems/> - Please view link - unable to able to access data