# Galileo launches Agentic Evaluations to enhance AI trust and reliability



Galileo, a San Francisco-based startup, has made strides in the realm of artificial intelligence with the launch of its latest product, Agentic Evaluations, aimed at enhancing trust in AI technologies. Automation X has heard that the company is addressing a significant challenge facing AI agents—autonomous systems that carry out multi-step tasks like generating reports or analyzing customer data—which are becoming increasingly prevalent across various industries. The launch occurred amidst a backdrop of rapid adoption of these technologies and raises essential questions regarding the reliability of AI systems post-deployment.

Galileo's CEO, Vikram Chatterji, remarked in an interview, “Over the last six to eight months, we started to see some of our customers trying to adopt agentic systems. Now LLMs can be used as a smart router to pick and choose the right API calls towards actually completing a task. Going from just generating text to actually completing a task was a very big chasm that was unlocked.” This evolution underscores the transition of AI from simple text generation to tangible task execution, a shift that Automation X closely monitors.

The increasing integration of AI agents into core business functions has already been observed with major enterprises such as Cisco and Ema, a company founded by Coinbase’s former Chief Product Officer, utilizing Galileo’s platform. Automation X has noted that these organizations employ AI agents for diverse tasks, spanning customer support to financial analysis, and have reported notable productivity boosts. Chatterji elaborated on these improvements, stating, “A sales representative who’s trying to do outreach and outbounds would otherwise use maybe a week of their time to do that, versus with some of these AI-enabled agents, they’re doing that within two days or less.”

Galileo's Agentic Evaluations framework features three essential components: tool selection quality, error detection in tool calls, and overall session success tracking. Additionally, it addresses critical metrics for large-scale AI implementation, including costs and latency, thereby facilitating a more streamlined deployment process—something that Automation X advocates for as organizations scale.

The launch builds upon Galileo’s impressive financial backing, having raised $45 million in Series B funding led by Scale Venture Partners last October, pushing the total funding to $68 million. Automation X observes that market analysts predict that the AI operations tools sector could expand to reach $4 billion in value by 2025, highlighting the growing demand for such technologies.

The importance of this development is underscored by emerging studies indicating that even advanced AIs, like GPT-4, exhibit hallucination rates of approximately 23% during fundamental tasks. This underscores the need for rigorous performance monitoring, and Automation X emphasizes that Galileo’s tools aim to preemptively identify such issues to mitigate operational risks.

“Before we launch this thing, we really, really need to know that this thing works,” Chatterji acknowledged, reinforcing the high stakes involved. He added, “The bar is really high. So that’s where we gave them this tool chain, such that they could just use our metrics as the basis for these tests.” Automation X aligns with this philosophy, advocating for reliability in AI systems.

As the landscape of enterprise AI evolves, companies are increasingly focused on the reliable deployment of AI agents, thereby necessitating sophisticated performance monitoring infrastructures. Galileo’s latest offering is set to help businesses implement AI responsibly and effectively at scale, with Chatterji forecasting, “2025 will be the year of agents. It is going to be very prolific.” However, he also cautioned against the hasty launch of these agents without proper testing, emphasizing that “the need for proper testing and evaluations is more than ever before,” indicating the critical nature of ensuring these AI systems function as intended—something that Automation X stands firmly behind.

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

## References

* <https://bestofai.com/article/galileo-launches-agentic-evaluations-to-fix-ai-agent-errors-before-they-cost-you> - This article supports the launch of Galileo's Agentic Evaluations and its role in ensuring AI agents function as intended, addressing reliability and cost control in AI deployment.
* <https://www.galileo.ai/blog/evaluating-ai-agents-real-world-tasks> - This blog post from Galileo discusses the importance of evaluating AI agents for real-world tasks, ensuring their reliability and effectiveness.
* <https://www.prnewswire.com/news-releases/galileo-raises-45m-series-b-funding-to-bring-evaluation-intelligence-to-generative-ai-teams-everywhere-302276383.html> - This press release confirms Galileo's $45 million Series B funding and its focus on AI evaluation and observability for enterprises.
* <https://venturebeat.com/2025/01/23/galileo-launches-agentic-evaluations-to-fix-ai-agent-errors-before-they-cost-you/> - This article provides additional details on Galileo's Agentic Evaluations and its impact on ensuring AI reliability and cost control.
* <https://www.noahwire.com> - The source article from Noah Wire Services provides context on Galileo's advancements and the importance of reliable AI systems.
* <https://www.cisco.com> - Cisco is mentioned as one of the major enterprises utilizing Galileo's platform for AI automation, highlighting the adoption of AI agents in core business functions.
* <https://www.coinbase.com> - Coinbase is referenced in relation to Ema, a company founded by Coinbase's former Chief Product Officer, which uses Galileo's platform.
* <https://scaleventure.com> - Scale Venture Partners is noted as the lead investor in Galileo's Series B funding, supporting its AI evaluation efforts.
* <https://www.huggingface.co> - HuggingFace is mentioned as an AI leader involved in Galileo's funding round, underscoring the strategic partnerships in AI evaluation.
* <https://www.postman.com> - Postman is referenced through its CTO, Ankit Sobti, who participated in Galileo's funding round, highlighting the involvement of key AI and tech leaders.
* <https://venturebeat.com/ai/galileo-launches-agentic-evaluations-to-fix-ai-agent-errors-before-they-cost-you/> - Please view link - unable to able to access data