Flash | Citi Research

11 Jun 2026 15:09:57 ET | 9 pages

Artificial Intelligence

At the Edge: Verification, Agents, and the Infrastructure Around Them

CITI'S TAKE
We attended two events this week that together map the frontier of applied AI: one on the infrastructure and deployment layer (Inflection), the other on the trust and reliability problem that becomes more urgent the more AI gets deployed (The Verification Summit). The industry is now wrestling with harder, less glamorous problems including orchestration, determinism, cost governance, and what it actually means to put an agent in production. Formal verification's "prove before deploy" model could be the key to scaling agentic AI adoption, while displacing a layer of observe-and-catch tooling (testing, auditing, LLM guardrails) that exists today precisely because outputs cannot be guaranteed.

Heath Terry[AC]
Shelby Spencer
Ashley Kim
Janna Withrow

Formal verification is emerging as a structural constraint on enterprise AI adoption, and an early-stage category worth watching. LLM failure modes are now visible enough that patches are accumulating faster than trust. Vinod Khosla noted that systems built on standard LLMs that look better than they are will not hold up over time. Formal verification encodes domain knowledge into machine-checkable logic so AI outputs can be traced and proved, not just evaluated probabilistically. Event host Pramaana Labs is "building the verification layer for AI" targeting real-world business domains like tax, legal, and healthcare, with a US tax offering expected in the next few months. Harmonic, Axiom, and Theorem are commercializing formal proofs via APIs, targeting software and hardware verification in institutions and enterprises where the cost of errors is high. Logical Intelligence is taking an architectural bet on energy-based models for combinatorial reasoning tasks where standard LLM structures fall short. Larger players are also active: Microsoft has deployed formal methods across Azure and low-level OS code; DeepMind is working on run-time verification for privacy and security in agentic systems. On the demand side, a hardware company representative expressed a strong desire to partner with formal verification vendors, as probabilistic outputs are not acceptable for objectives such as GAAP compliance. We see formal verification as a meaningful accelerant for enterprise agentic AI adoption, by addressing the need for guaranteeing deterministic outputs in high-stakes domains.

Most organizations have prototyped agents. Very few have productionalized them. Cognition noted that startups are on top of agentic AI while corporates may be months or years out. Guardrails AI described the "infinite surface area problem," where unlike traditional software, you cannot enumerate all the ways an agent might fail before deployment. Their solution is pre-production simulation, running synthetic queries at scale to identify failure modes, with the view that orchestration and evaluation need to be coupled rather than handled by separate products. Pinecone noted that in September 2025, agents surpassed humans as the largest class of API callers on their platform. With Pinecone Nexus, they are rearchitecting the knowledge retrieval layer for agents, claiming 91–95% token reduction versus human-oriented retrieval architectures.

Token costs are a governance and measurement problem, and model routing is becoming a real capability for more sophisticated operators. Databricks confirmed that internally, AI budgets for engineers are effectively unlimited, with most spend on

See Appendix A-1 for Analyst Certification, Important Disclosures and Research Analyst Affiliati

a leading coding platform. Customers, however, want cost controls, and Databricks' AI Unity Gateway provides visibility on utilization and cost so enterprises can see where spend is going. Cognition sees the real question as how to use cheaper models for 50–60% of tasks that do not require frontier capability. DataRobot noted that enterprises are already moving between models depending on use-case.

Physical AI safety is formalizing, and the toolchain is open-sourcing faster than the market appreciates. UC Berkeley's Sanjit Seshia discussed progress in verification in physical/embodied AI, with his Scenic probabilistic programming language and VerifAI back-end tool already deployed in AV, humanoid, and aerospace contexts. NVIDIA's Marco Pavone discussed the Alpamayo open platform, comprising open-weight vision models, an end-to-end open simulator (AlpaSim), and physical AI datasets. The broader implication is that the formal verification costs are declining, and formal methods need to be embedded across the full design loop rather than applied as a single end-of-pipeline check.