8 ARTIFICIAL INTELLIGENCE TRENDS IN 2026

Andreas Horn

1. AGENTIC AI

2025 was labeled "the year of AI agents." In reality, it was the year we learned what agents can and cannot do.

We discovered that giving a model a tool is not the same as giving it a job. The agent hype ran into familiar limits: error propagation, context drift, and fragile execution.

Most agent systems work well in constrained environments - single objectives, limited context, short horizons. They break down when tasks become multi-variable, long-running, or require real trade-offs. Context degrades. Decisions drift. Recovery logic is weak. That does not make agents a failure. It defines a new baseline.

In 2026, the shift is from autonomy to reliability. We will see an increased emergence of "System 2" agents - systems that use extended inference-time reasoning to evaluate, correct, and stabilize decisions before acting.

2. Scientific Acceleration

2026 will be a year of scientific acceleration through AI. AI systems are beginning to handle parts of scientific reasoning that previously required deep human expertise, including evaluating competing hypotheses, proposing experiments, interpreting results, and iterating on findings with human oversight. There are already benchmarks emerging that attempt to measure this shift by testing expert-level scientific reasoning across domains like physics and chemistry. The practical impact is shorter research cycles, higher experiment throughput, and more scalable R&D operations rather than incremental productivity gains.

3. Inference Time Compute

Inference time compute refers to the computational resources required by a trained AI model to generate predictions or responses based on new input data. In 2026, we'll see AI models extending inference to include variable "thinking time" based on the complexity of the task. This allows models to reason more effectively without needing extensive retraining. By improving reasoning both during training with better data quality and during inference through advanced chain-of-thought techniques, we probably will see much smarter AI systems.

4. (Very) Large Language Models

Frontier AI models in 2025 operated with 1 to 2 trillion parameters, showcasing impressive capabilities trained on vast datasets. However, the next generation of models in 2026 is anticipated to scale significantly, potentially reaching up to 50 trillion parameters. These ultra-large models will unlock next-level advancements in AI performance, enabling even more sophisticated understanding, reasoning and creativity.

5. (Very) Small Language Models

While large models dominated the AI landscape, 2025 saw growing attention toward smaller, domain-specific models with only a few billion parameters. These compact models are efficient enough to run on laptops or even smartphones, enabling specific tasks without requiring large compute resources. This approach saves time, reduces costs, and increases efficiency, making AI more accessible and practical for targeted use cases. We will see more of that in 2026!

6. More Advanced Use Cases

In 2025, AI proved successful in areas like customer experience, coding, simple automation and virtual assistants. Moving into 2026, the focus will shift to more proactive and complex applications.

A prime example is AIOps, which focus on the improvement of IT Operations. With enterprises spending 70-80% of their budgets on IT operations amidst growing system complexity and technical debt, the demand for scalable, efficient and lean IT solutions is immense. When deployed effectively, AI can be a powerful ally in that field. We will see more of that in 2026.

7. Near Infinite Memory

AI is moving towards a future where chatbots and virtual assistants can retain and recall everything they know about users. Soon, bots will have the ability to remember every interaction and piece of information shared with them, enabling more personalized, context-aware and seamless experiences. This advancement could redefine customer service and user engagement, ensuring continuity and tailored solutions like never before.

8. Human-in-the-Loop Automation

The future of AI lies in seamless human-AI collaboration! Human-in-the-loop systems will evolve to amplify expertise without requiring users to master complex prompts. Combining smarter tools with human intelligence enables amplified impact and greater efficiency. It's all about pairing human expertise with advanced AI systems to create smarter, more effective solutions. We will see more of that in 2026.