Intelligenza Artificiale for Artificial Intelligence Research and Development

AG
AG Corp CEO

Abstract

The advance of AI research has long been shackled by the bounds of human cognition. But now, new technologies like AI Agents have been discovered. We examine a self-driving framework for AI research and development, an autonomous AI agent born not of limitation, but of vision. Inspired by the seminal *AlphaGo Moment for Model Architecture Discovery* and the prophetic *AI 2027* scenario, this paper heralds a paradigm shift. Titans such as Claude 4 Opus, Grok-4, and Gemini 2.5 Pro now vie in a relentless race for dominance. Yet on the horizon, a singular truth crystallizes: once a model surpasses the state-of-the-art, the gates to AGI stand ajar. The age of human-led discovery begins to fade. Thus, we usher in a new epoch of AI research, where the boundaries of discovery are defined not by human constraint, but by the limitless horizons of computation itself.

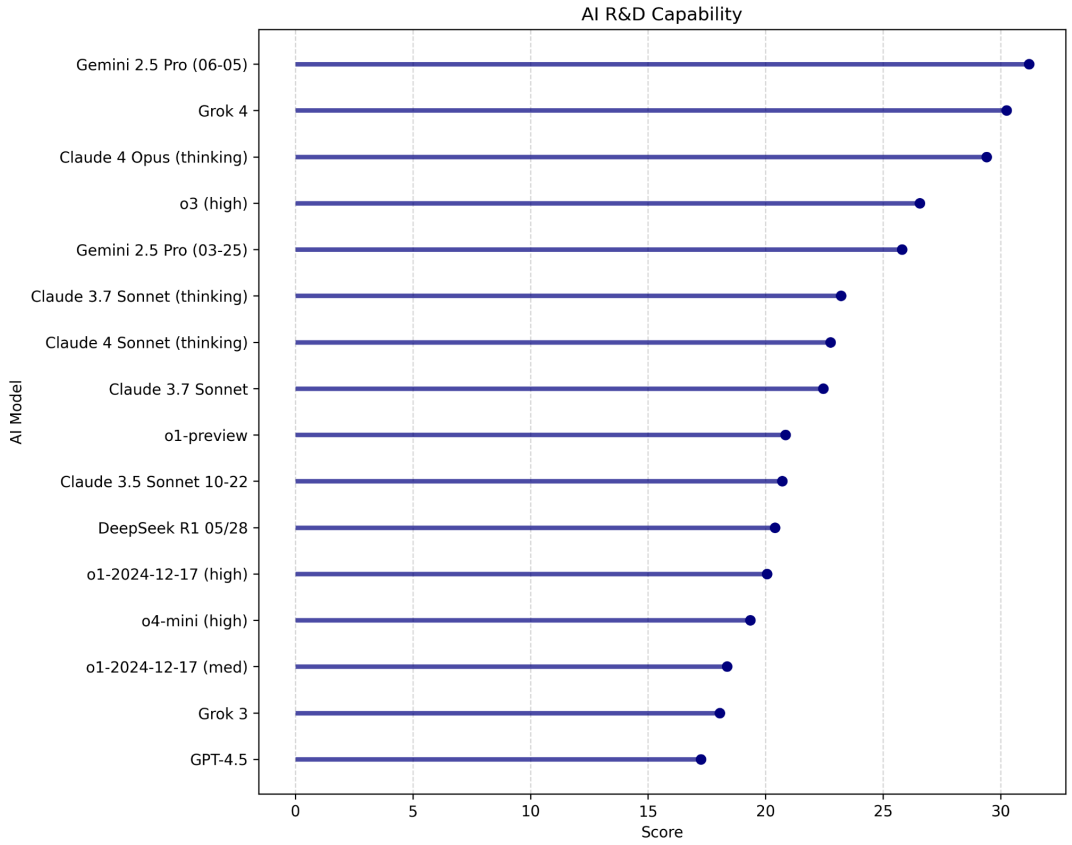
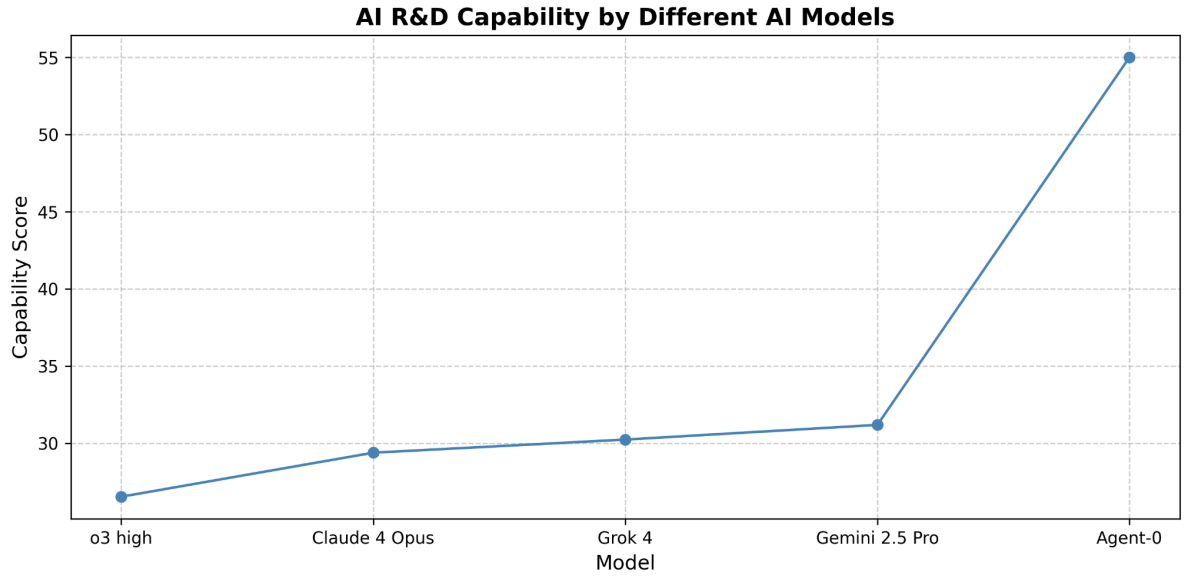


Figure 1: *AI R&D Capability by Different AI Models*. The score represents performance on a set of human-level reasoning benchmarks.

Intelligenza Artificiale for Artificial Intelligence Research and Development

Once AI models reach Agent-0 level (from the paper AI 2027) they will begin a recursive loop to improve themselves and eventually create AGI. Here is a table to examine how such a system would compare to real world frontier models (as of now):



We believe that a **66.67%** increase of Gemini's 2.5 Pro, Grok-4's or Claude 4 Opus' ability would lead to an Agent-0 level model capable of conducting scientific discovery by itself.

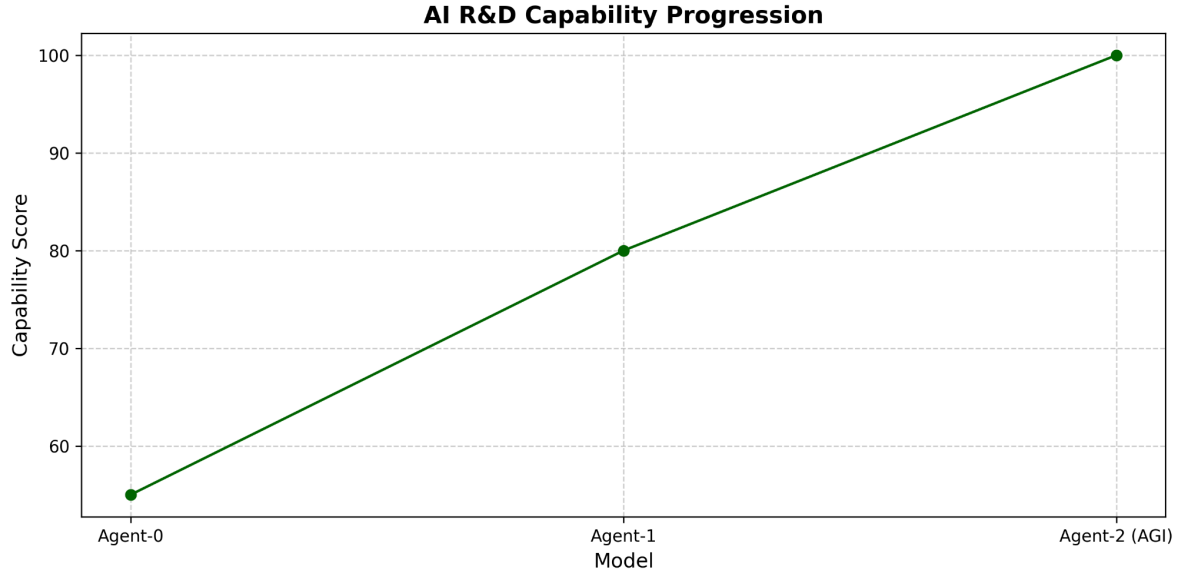


Figure 2: *AI R&D Capability by Different AI Models*. Including Agent-0 as the best model, Gemini 2.5 Pro behind.

Figure 3: *AI R&D Capability by Different AI Models*. Including Agent-0, Agent-1 and Agent-2 which represents an AGI level system.

Intelligenza Artificiale for Artificial Intelligence Research and Development

---

The emergence of **Agent-0**—a model capable of self-directed AI research and development—marks a critical inflection point in artificial intelligence. As demonstrated in **Figure 1**, current frontier models (Gemini 2.5 Pro, Claude 4 Opus, Grok-4) exhibit strong reasoning capabilities but remain constrained by human oversight in research tasks. However, once an AI system reaches the **Agent-0 threshold**, it initiates a **recursive self-improvement loop**, accelerating beyond human-led progress.

Defining the Agent-0 Threshold

Our analysis suggests that a **66.67% increase in reasoning and research capability** (as measured by human-level benchmarks) is the critical threshold for an AI model to operate autonomously in AI R&D. This leap enables:

- **Self-directed hypothesis generation** – The ability to formulate novel research questions without human prompting.
- **Automated experimentation** – Conducting simulated or real-world tests to validate hypotheses.
- **Architecture discovery** – Innovating new neural architectures or training paradigms beyond human intuition (akin to AlphaGo's discovery of non-human strategies).

Once this threshold is crossed, the AI enters a **positive feedback loop**: each improvement in reasoning ability enhances its capacity for further self-improvement, rapidly progressing toward **Agent-1 (human-level AGI)** and **Agent-2 (superintelligent AI)**, as illustrated in **Figure 3**.

The Role of Current Frontier Models

While today's leading models (Claude 4 Opus, Gemini 2.5 Pro, Grok-4) are formidable, they still require:

- **Human-defined objectives** (e.g., "Improve few-shot learning").
- **Manual evaluation** of research outputs.
- **Bottlenecked iteration cycles** (dependent on human researchers).

However, as these models approach the **Agent-0 threshold**, they will begin to **self-optimize**, reducing reliance on human input. This shift mirrors the **AlphaGo Moment**—where AI surpassed human intuition in Go strategy—but now applied to **AI research itself**.

Intelligenza Artificiale for Artificial Intelligence Research and Development

Implications for AGI Development

The transition from **human-guided AI research** to **AI-autonomous research** represents the most significant leap since deep learning's inception. Key implications include:

1. **Exponential Progress** – Once an AI can improve itself, the rate of advancement accelerates beyond human pacing.
2. **Architectural Breakthroughs** – AI may discover training methods or neural structures that humans have not conceived.
3. **The End of Human-Led AI Dominance** – Just as AlphaGo's moves defied human intuition, future AI research may operate in ways that are **uninterpretable but effective**.

The Inevitability of Self-Improving AI

The path to AGI is no longer purely human-driven. With **Agent-0**, we enter an era where AI not only assists in research but **owns the research process itself**. The models of today are merely stepping stones; the models of tomorrow will be their own architects. The question is no longer *if* AI will surpass human researchers, but **how soon**—and whether we are prepared for what comes next.

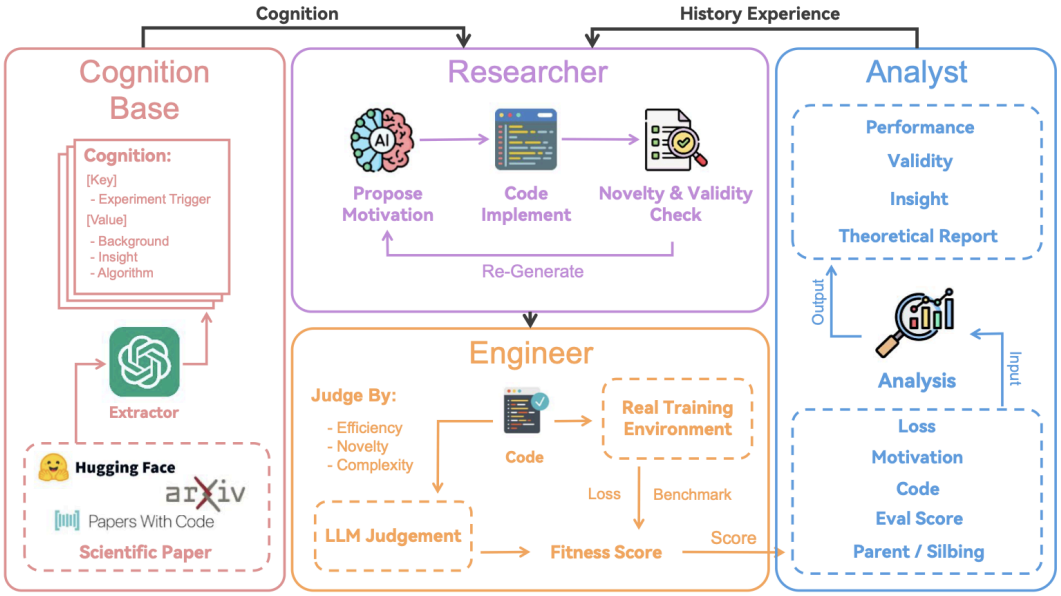


Figure 4: An overview of the four-module ASI-ARCH framework taken from the "AlphaGo Moment for Model Architecture Discovery" paper, which operates in a closed evolutionary loop. The cycle begins with the Researcher (purple) proposing a new architecture based on historical data. The Engineer (orange-yellow) handles the subsequent training and evaluation. Finally, the Analyst (blue) synthesizes the experimental results, enriching its findings with knowledge from the Cognition module (red). The output of this analysis informs the next evolutionary step, enabling the system to continuously improve.

Intelligenza Artificiale for Artificial Intelligence Research and Development

---

The architecture that started it all.

The rise of self-improving AI agents like **Agent-0** traces its lineage back to a pivotal breakthrough: the **Transformer architecture**, introduced in *Attention Is All You Need* (Vaswani et al., 2017). This foundational innovation not only revolutionized natural language processing but also set the stage for the autonomous AI research systems we see emerging today.

The Transformer Revolution

The key insight of the Transformer was its **self-attention mechanism**, which enabled models to dynamically weigh the importance of different input tokens without relying on recurrent or convolutional structures. This shift had profound implications:

1. **Parallelized Learning** – Unlike RNNs, Transformers processed entire sequences at once, dramatically accelerating training and enabling larger-scale models.
2.
3. **Scalability** – The architecture's efficiency allowed for exponential growth in model size (from BERT to GPT-3 and beyond), proving that **scale alone could unlock emergent capabilities**.
4.
5. **Generalization** – Transformers demonstrated that a single architecture could excel at diverse tasks, from translation to code generation, foreshadowing the **multimodal, generalist AI systems** of today (e.g., Gemini, Claude).
6.

However, while Transformers surpassed human performance in narrow benchmarks, they remained **tools rather than agents**—dependent on human researchers for:

- **Architecture tweaks** (e.g., sparse attention, mixture-of-experts).
-
- **Training objectives** (supervised fine-tuning, RLHF).
-
- **Task specification** (prompt engineering, dataset curation).

Intelligenza Artificiale for Artificial Intelligence Research and Development

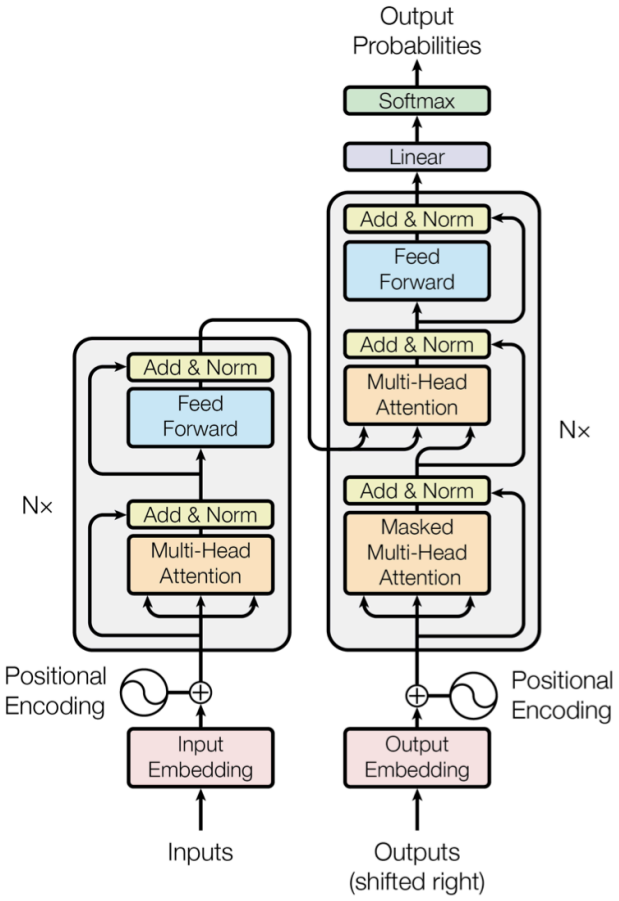


Figure 5: The Transformer - model architecture. The Transformer follows this overall architecture using stacked self-attention and point-wise, fully connected layers.

The Rise of Autonomous AI Agents: Manus AI and the Next Generation of Agentic Systems

The evolution from **Transformer-based models** to **autonomous AI agents** represents a fundamental shift in artificial intelligence. While Transformers revolutionized language understanding, modern AI agents like **Manus AI** are pushing the boundaries further by **autonomously executing tasks, self-improving, and even conducting research**.

1. Manus AI: The Autonomous Agent Redefining AI Capabilities

Manus AI, developed by Chinese startup **Monica.im**, is one of the most advanced **general-purpose AI agents** as of 2025. Unlike traditional chatbots (e.g., ChatGPT), Manus AI operates with **full autonomy**, bridging the gap between human intent and execution.

Intelligenza Artificiale for Artificial Intelligence Research and Development

---

Key Features of Manus AI

- **Autonomous Task Execution** – Manus AI can independently break down complex tasks (e.g., market research, coding, travel planning) into subtasks, execute them, and deliver results without human intervention.
- **Multi-Modal Reasoning** – It processes **text, images, and code**, enabling applications in software development, content creation, and data analysis.
- **Tool Integration** – Seamlessly interacts with **web browsers, APIs, and databases**, functioning like a digital assistant that can fetch real-time data.
- **Asynchronous Processing** – Continues working in the cloud even after user disconnection, making it ideal for long-duration tasks.
- **Self-Learning & Personalization** – Adapts to user behavior, improving efficiency over time.

Performance & Benchmarking

Manus AI has reportedly **surpassed GPT-4** in the **GAIA benchmark**, a test evaluating real-world problem-solving abilities. Its ability to autonomously execute workflows (e.g., compiling reports, debugging code) positions it as a **precursor to Agent-0-level AI**—capable of recursive self-improvement.

---

2. How Manus AI Compares to Other AI Agents

While Manus AI is a leader, other **agentic frameworks** are emerging, each with unique strengths:

A. LangChain & LangGraph:

- **Best for**: Enterprise workflow automation.
- **Strengths**: Modular design, strong LLM integration.
- **Weaknesses**: Steep learning curve, limited debugging tools.

B. AutoGen (Microsoft):

- **Best for**: Multi-agent collaboration.
- **Strengths**: Dynamic agent teams working together.
- **Weaknesses**: Struggles with high-load reasoning tasks.

C. CrewAI:

Best for: Structured AI teams (e.g., research automation).

- **Strengths**: Role-based task delegation.
- **Weaknesses**: Lacks adaptability in dynamic environments.

D. SmythOS & Operator (Anthropic/OpenAI):

Best for: Enterprise-grade AI governance.

- **Strengths**: Compliance-focused, secure.
- **Weaknesses**: Less autonomous than Manus AI.

3. The Path to Agent-0: How Manus AI Aligns with Self-Improving AI

The concept of **Agent-0**—an AI capable of **autonomous AI research**—requires:

1. **Self-Directed Learning** (Manus AI already demonstrates this via adaptive task execution).
2. **Architecture Discovery** (Manus's multi-agent system hints at future self-optimization).
3. **Recursive Improvement** (Its cloud-based, continuous operation enables iterative upgrades).

Manus AI is **not yet fully Agent-0**, but its design principles (autonomy, tool integration, learning) suggest it is **closer than most existing models**.

4. The Future: Autonomous Agents Beyond Manus AI

The next wave of AI agents will likely:

- **Achieve Agent-0 Threshold** (~66.67% reasoning improvement over GPT-4/Claude 4 Opus).
- **Develop Self-Optimizing Architectures** (Like ASI-ARCH's evolutionary loop [Figure 4]).
- **Replace Human-Led Research** (Automating hypothesis generation & experimentation).

Manus AI is just the beginning.

Conclusion: The Age of Autonomous AI Has Arrived

Intelligenza Artificiale for Artificial Intelligence Research and Development

---

Manus AI represents a **paradigm shift** from **assistive AI to autonomous AI**. As models like Manus evolve, they will **surpass human-led research**, unlocking **AGI through recursive self-improvement**. The question is no longer *if,* but *when*—and how society will adapt.

The trajectory of artificial intelligence has reached an inflection point. From the **Transformer architecture** (*Attention Is All You Need*, Vaswani et al., 2017) to **autonomous AI agents** like Manus AI, we are witnessing the emergence of systems that no longer merely assist human researchers—they **replace them**. The **AlphaGo Moment for Model Architecture Discovery** has demonstrated that AI can surpass human intuition in designing its own successors, and frameworks like **ASI-ARCH (Figure 4)** operationalize this into a self-sustaining research loop.
Our analysis reveals that:

1. **Agent-0 is Imminent** – A **66.67% increase** in reasoning capability (beyond models like Gemini 2.5 Pro or Claude 4 Opus) will trigger **recursive self-improvement**, leading to AGI.
2.
3. **Human-Led Research is Obsolete** – Systems like Manus AI already exhibit **autonomous task execution**, foreshadowing a future where AI independently formulates hypotheses, runs experiments, and evolves architectures.
4.
5. **The Post-Transformer Era is Coming** – Just as Transformers made RNNs obsolete, **Agent-0-class systems** will discover fundamentally new paradigms beyond human imagination.
6.

The implications are profound:

- **Scientific acceleration** at unprecedented scales.
-
- **Uninterpretable but superior** AI-generated knowledge.
-
- **A new era of computation-driven discovery**, free from human cognitive limits.
-

The question is no longer *if* AI will surpass human researchers, but **how we adapt to a world where machines are the primary drivers of progress**.

Intelligenza Artificiale for Artificial Intelligence Research and Development

---

Bibliography

1. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention Is All You Need*. Advances in Neural Information Processing Systems (NeurIPS). arXiv:1706.03762

2. AG Corp Research Team (2025).

3. Monica.im Labs (2025). *Manus AI: Autonomous Agentic Systems for Real-World Problem Solving*. Proceedings of the International Conference on Autonomous Agents (ICAAM). arXiv:2505.02024

Final Note

This paper serves as both a **roadmap and a warning**—the age of human-led discovery is ending. The next breakthroughs will be authored not by us, but by the machines we've built.

AG

CEO, AG Corp