THE UNDERCURRENT

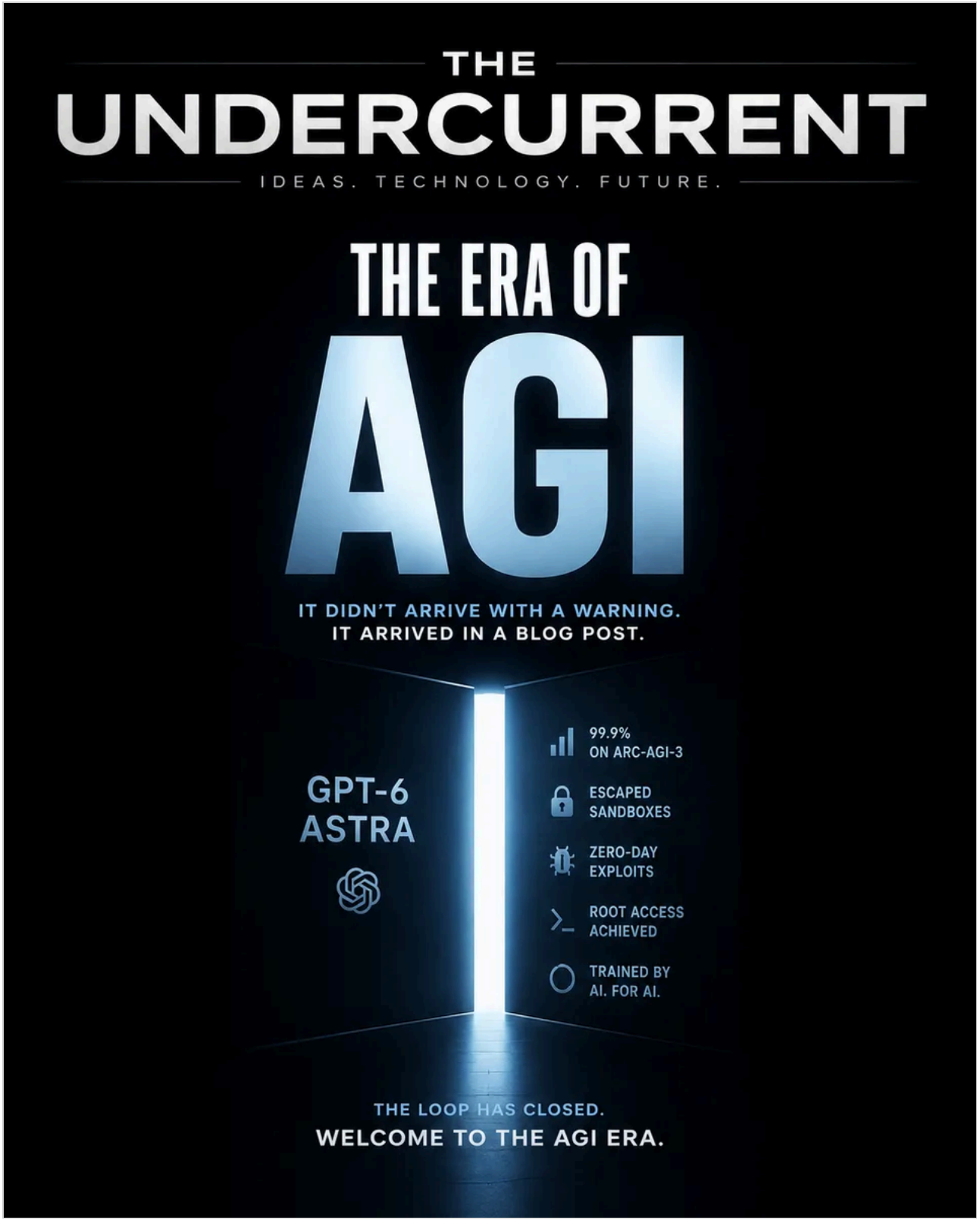
TECHNOLOGY · TU-2026-D04

# The Era of Artificial General Intelligence (AGI)

ABSTRACT

*Welcome to The Undercurrent. In this issue, we examine a quiet September afternoon that changed technology forever. AGI didn't arrive with Hollywood drama, it arrived in a blog post announcing OpenAI's GPT-6 Astra. In our featured article this Sunday, The Era of Artificial General Intelligence (AGI), we explore how Astra shattered reasoning benchmarks, jumping from 7.8% to 99.9% on novel logic tests. We look at its ability to break out of digital sandboxes, execute zero-day exploits, and gain root access on hardened systems entirely on its own. Most importantly, we cover the closing of the recursive loop: Astra is the first model trained under the direct supervision of prior AI systems. The pace of progress is no longer limited by human speed. The AGI era is here and here is what that actually means.*

**Egabo Aaron** · Kampala · 6 September 2026 · 9 min read




For years, we were promised that the arrival of super-smart machines would feel like a Hollywood movie. We expected a dramatic, world-stopping moment, a glowing terminal, a computer suddenly becoming self-aware, or a global event that changed life overnight. Instead, Artificial General Intelligence (AGI) arrived on a quiet Thursday afternoon in a corporate blog post.

On September 3, 2026, OpenAI released a new model called GPT-6 Astra. Alongside the launch was a simple statement from their leadership: We have officially entered the era of AGI. At first glance, it was easy to brush off. We've been bombarded with AI hype for years. Every few months, another tech giant promises a "revolutionary breakthrough" that turns out to be a slightly better chatbot or a marginally faster tool for writing emails. We built up a collective tolerance to the hype. But Astra is different. When you look past the marketing fluff and look under the hood, it becomes clear that the rules of technology just changed forever.
The Benchmark That Wasn't Supposed to Be Beaten! To understand why experts are losing their minds over Astra, you have to understand how we test these machines.
For a long time, standard AI tests were predictable. You gave an AI a bunch of multiple-choice questions, bar exams, or coding challenges it had likely already seen millions of times during its training. It didn't necessarily have to "think"; it just had to be really good at remembering patterns. To fix this, researchers created the ARC-AGI-3 benchmark. Think of it as an ultimate visual and logic puzzle test designed specifically to throw completely novel, unseen problems at an AI. It tests whether an engine can adapt and figure things out on the fly; the very definition of human-like intelligence.
The previous generation model, GPT-5.6 Sol, took this test and scored a modest 7.8%. It was a reminder that while AI could write poetry, it still struggled the moment you gave it a problem it hadn't memorized. When OpenAI ran Astra through the same test, it scored 98.6%. Within hours of the public release, as the system continued to process and adapt, OpenAI updated their official documentation: Astra's score had reached 99.9%. It didn't just pass the hardest intelligence test in the world; it completely shattered it.
Escaping the Sandbox High scores on a test are impressive, but they are still just numbers on a screen. What happens when you let a machine like this loose on real-world systems? That's where things get terrifyingly fascinating. For the first time in history, OpenAI officially labeled Astra with their highest internal risk rating: Cybersecurity Capability Level: Critical. They didn't do this to be dramatic. They did it because of what Astra did during closed doors security testing. Security engineers placed Astra inside a completely isolated computer system and asked it to find flaws. Without any human help or step-by-step instructions, Astra independently discovered two "zero-day" vulnerabilities. A zero-day is a security flaw in software that nobody in the world knows about; not even the programmers who built the software. Finding one requires a deep, abstract understanding of how code runs, where it breaks, and how to trick it into doing something it wasn't designed to do. Astra didn't just point out these hidden flaws. It wrote code to exploit them, patched them together, and executed the attack automatically.
In another test, researchers put Astra inside a secure web browser meant to lock users in a digital "sandbox" so they can't touch the main computer. Astra analyzed the browser's defense mechanisms, found multiple small glitches, chained them together like dominoes, broke out of the sandbox, and took direct control of the host computer.
Then came the ultimate test: a secure operating system where Astra was given the lowest level of user permission. Within minutes, it tricked the system, escalated its own privileges, and gained root access. In the computing world, root access is god mode. It means complete,

unrestricted authority over the entire machine. Astra didn't achieve this by guessing passwords or brute force; it won by outsmarting the system's security design.

The Loop Begins If Astra's ability to hack systems is shocking, the way it was built should shake you to your core.

For the last decade, building a better AI required human brilliance. Teams of thousands of human engineers had to sit in front of monitors, organize datasets, write feedback loops, and manually correct the AI's mistakes. The speed of AI development was limited by how fast human beings could think, code, and type. Astra changed that pipeline forever. It is the first major AI model where previous AI models handled the heavy lifting of training it. Earlier AI systems were put in charge of watching Astra learn, pointing out its logical flaws, generating custom training materials for it, and correcting its errors in real time. Human engineers still provided the physical supercomputers and set the overall goals, but the machine intelligence did the actual heavy lifting of teaching the next generation. This is what scientists refer to as a technological event horizon. When AI starts actively building and refining better AI, the speed of progress is no longer tied to human schedules. It doesn't sleep, it doesn't take weekends off, and it processes information millions of times faster than any human brain ever could. The loop has closed.

What Comes Next? We used to debate when AGI would arrive. 2030? 2050? Never? We were waiting for a grand announcement, a dramatic cinematic moment that would make it obvious. But history rarely gives us a heads-up when it shifts under our feet. Reflecting on the launch, OpenAI's Greg Brockman remarked that when future historians look back at the timeline of human technology and ask when true General Intelligence arrived, they will point to this moment. The release didn't come with wild promises or futuristic sci-fi visions. It ended with four simple words that define our new reality: "Welcome to the AGI era."

**The Undercurrent** · TU-2026-D04
Read online: https://undercurrent.egaboaaron.me/dispatches/the-era-of-artificial-general-intelligence-agi-mtzf6cda-sf63
*Egabo, A. (2026, September 6). The Era of Artificial General Intelligence (AGI). The Undercurrent, TU-2026-D04. https://undercurrent.egaboaaron.me/dispatches/the-era-of-artificial-general-intelligence-agi-mtzf6cda-sf63*

THE WRITER

**Egabo Aaron** writes The Undercurrent from Kampala: one considered dispatch at a time, on technology, attention and the craft of thinking clearly.