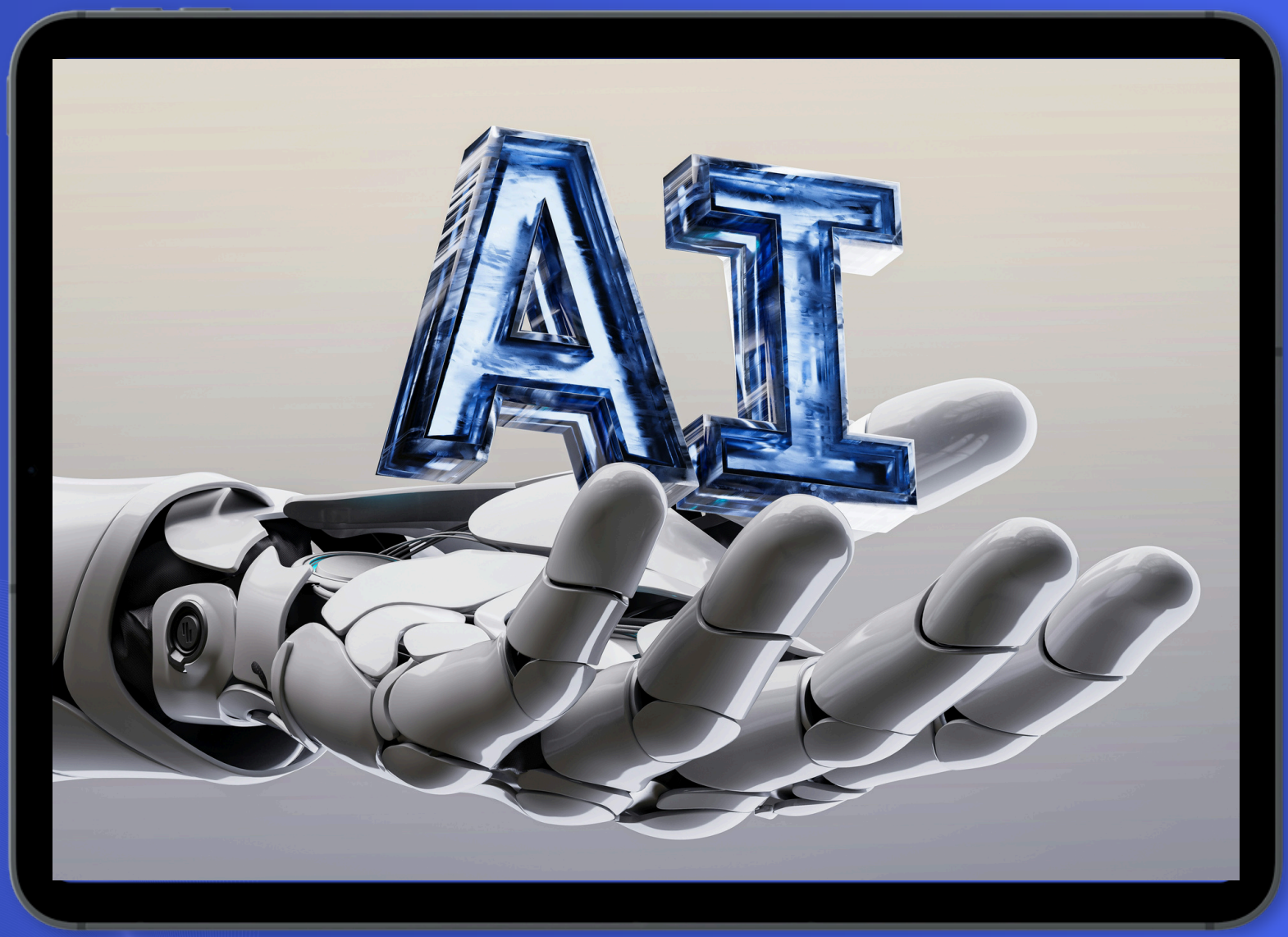
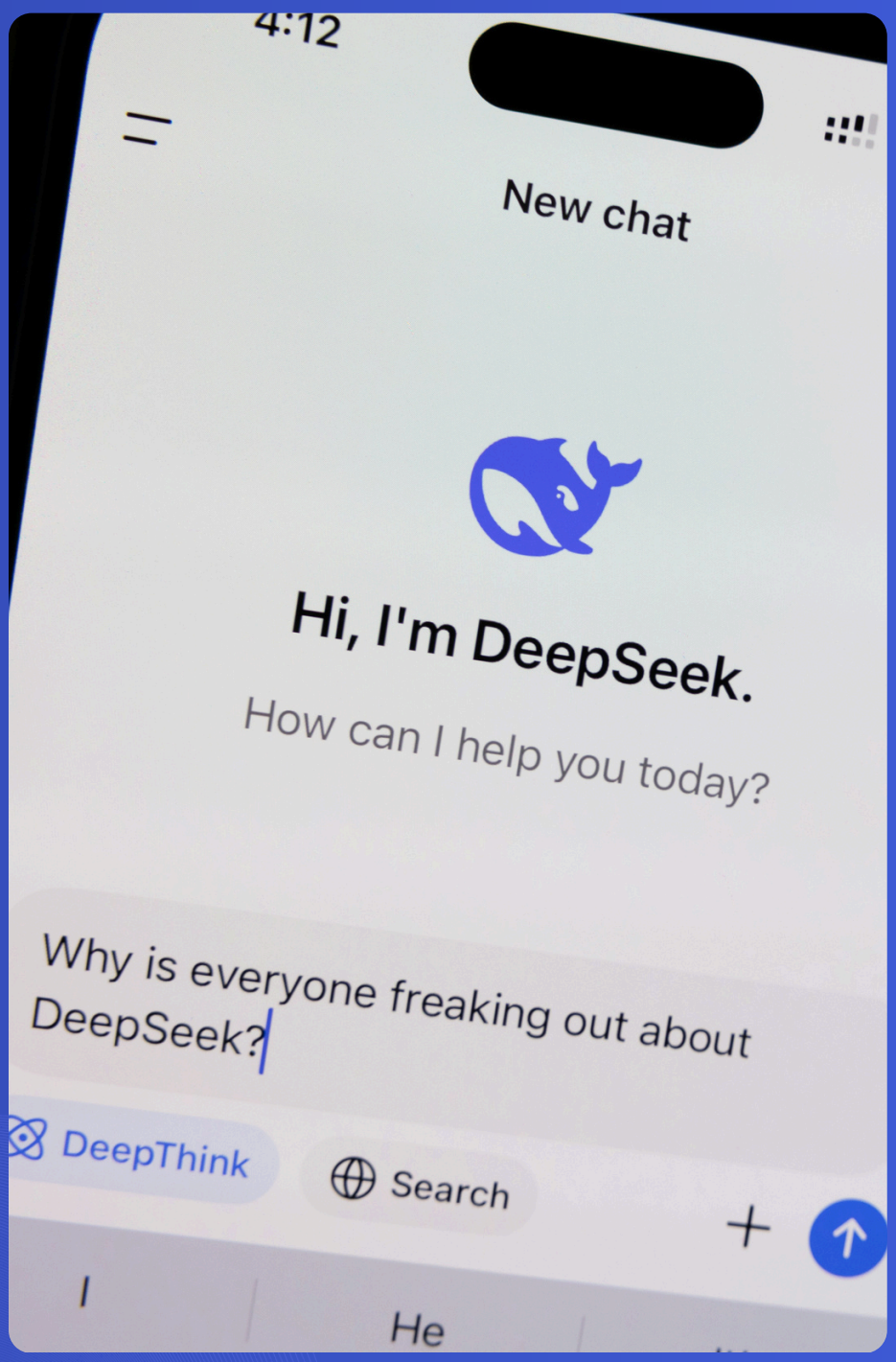
THE MAJOR AI TRENDS OF 2025



01 Comeback Mode

Google's resurgence is thanks to Gemini 3's superior performance and impressive imaging tools such as the Nano Banana Pro. It also launched Ironwood — its seventh-gen Tensor Processing Units (TPUs) designed for the “age of inference” and can be used by companies to train AI models. Google is now positioned as a potential threat to Nvidia's dominance on the chip market. Warren Buffett's Berkshire Hathaway is also betting on its prospects and has made its first-ever $4.3 billion investment in Alphabet.

02 China's AI Ambitions



In January, DeepSeek from China stunned the tech world by launching R1 — a platform that could rival OpenAI in benchmark tests but at a lesser cost and computing power. Since then, DeepSeek has gone quiet and the launch of R2 is awaited.

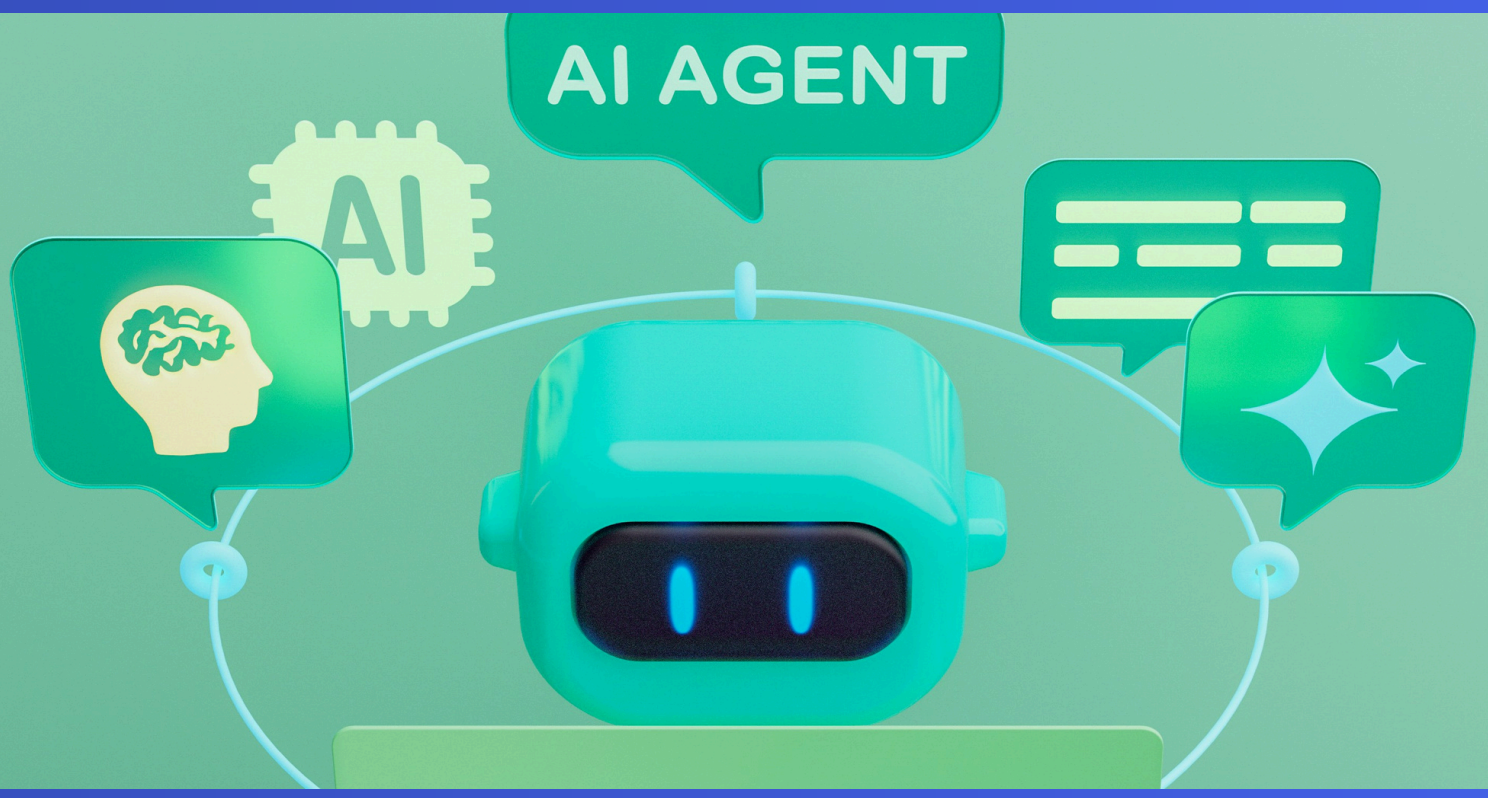
On the other hand, Alibaba has launched a powerful AI model, and Tencent released Hunyuan-A13B — an AI model designed to be faster, smarter and open to developers.

03 Vibe Design

In February, OpenAI cofounder Andrej Karpathy coined the term "vibe-coding" to describe AI prompted by natural language to assist with coding. Before year-end, it made it to Collins Dictionary as the Word of the Year.

While vibe coding can speed up work and even help novices design sites, the code can have bugs and vulnerabilities that could be costly in the long run to fix.

04 Agentic AI



These are a subset of Gen AI that can automate complex business processes. Instead of just responding to commands, it can plan, reason, and take action. While there was progress in 2025, it fell short of expectations.

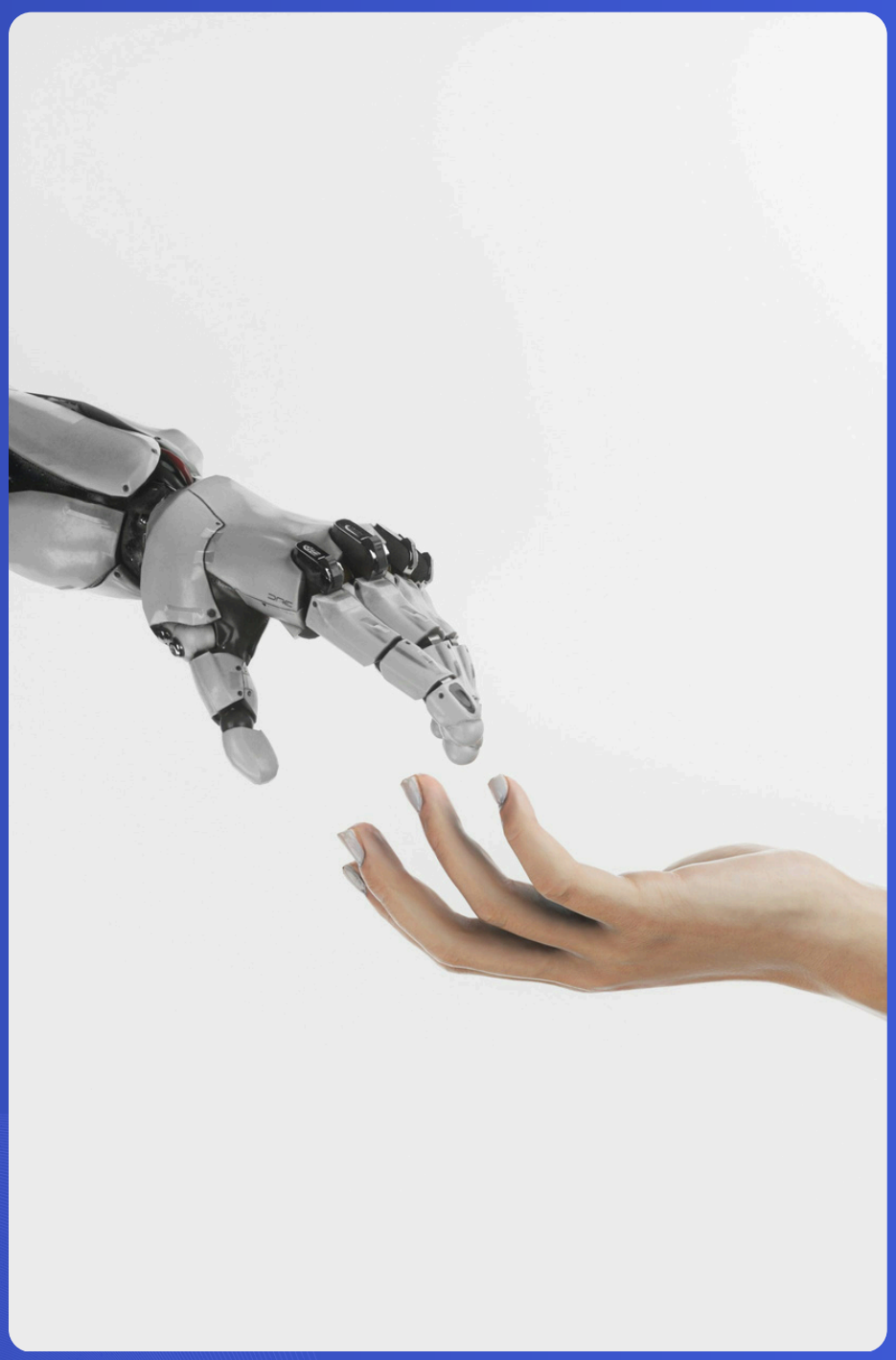
To unlock its revenue potential, we need to reimagine workflows with agents at the core, says a McKinsey report. Organizations need to go beyond siloed teams to cross-functional squads, upskill staff, and adapt tech infrastructure.

05 Promise, Hype & Reality

AI brings the promise of transforming operations, unlocking revenue and delivering a strategic advantage. But it's yet to translate into returns. As per a BCG report, three-quarters of companies (74%) are struggling to capture value from AI. Most companies need to scale AI and weave into the fabric of operations, the study said.

AI is also being used to justify layoffs. But, the truth is that at present, AI needs human interaction to operate.

06 Automation To Autonomy

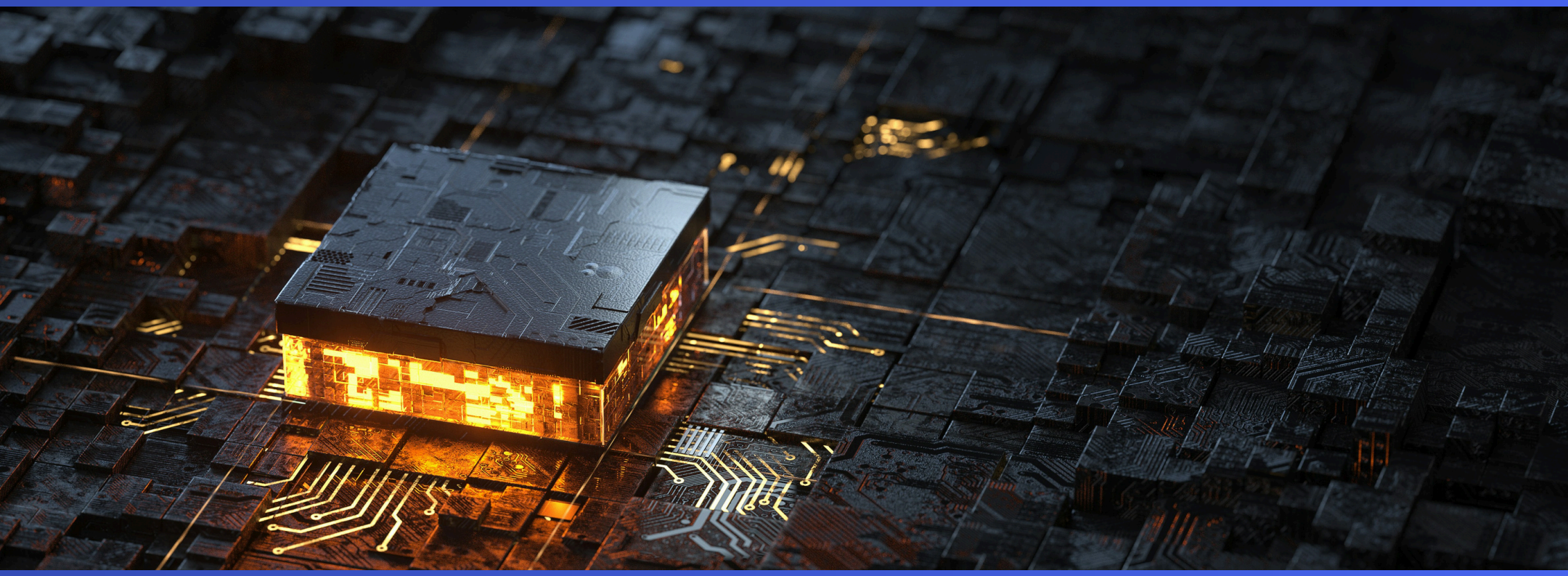


AI is transforming sectors via automation. It plays to its strengths of spotting patterns and doing repetitive tasks. Yet, autonomy is still a while away.

Instead, optimal results will come from human + AI collaboration where humans control critical decisions.

For instance, Agentic AI has agents handle key tasks but humans define the goals.

07 Chips Diversification



The AI chip war is heating up. What used to be a niche segment of the semiconductor industry now has big tech vying for gains.

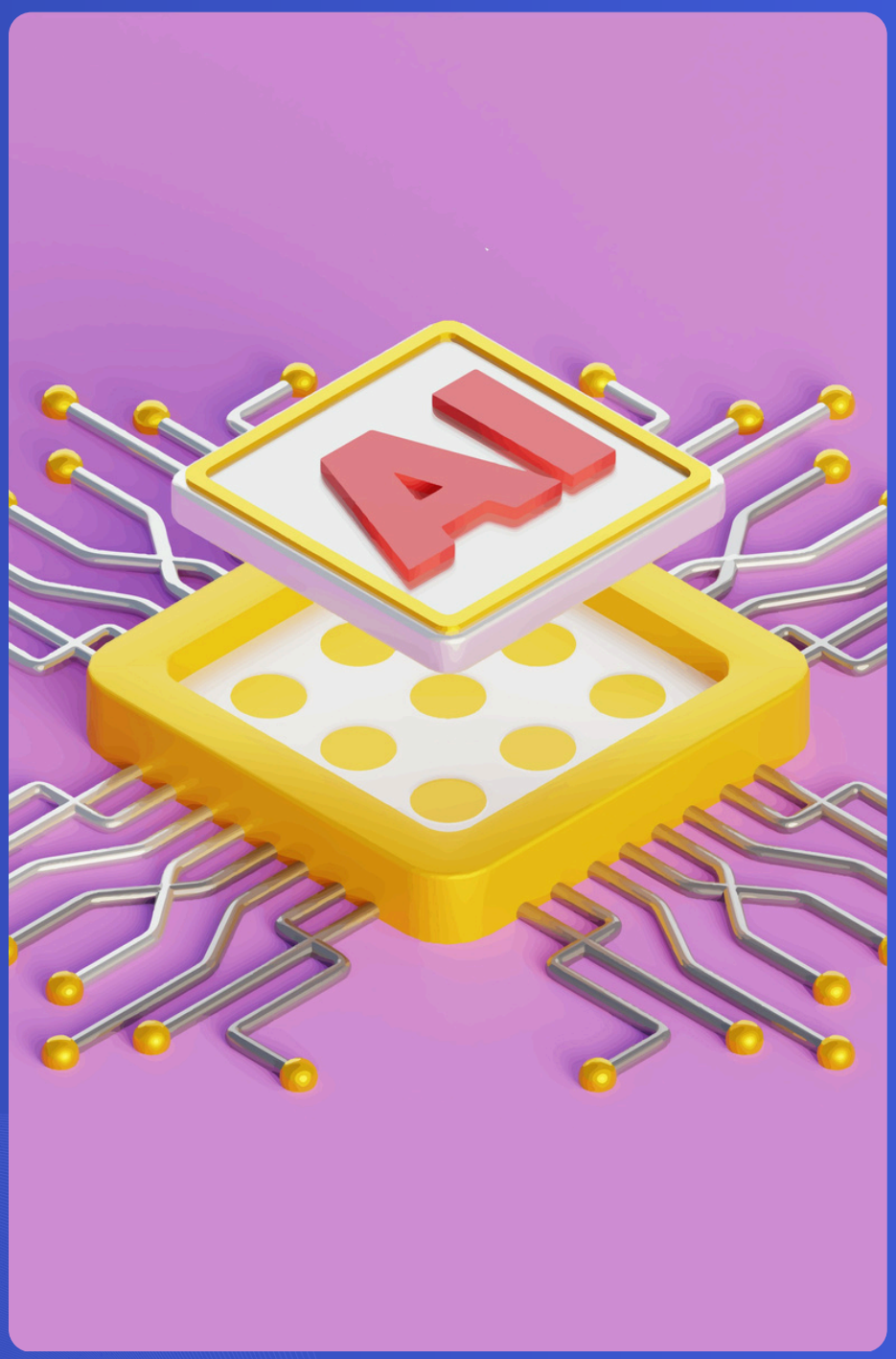
While Nvidia leads in Graphics Processing Units (GPUs) to accelerate computing, Google offers Tensor Processing Units (TPUs) which optimise ML tasks in data centres and is rented via Google Cloud. AMD offers diversified hardware, open-source software, and superior central processing units (CPUs). The quest is on for more energy-efficient chips with a focus on edge processing to perform inferential tasks.

08 Rise Of SLM

There's much talk about LLMs (large language models), but 2025 has put the spotlight on SLMs (small language models). These are lightweight, cheaper and solve specific problems in a lean and efficient manner.

They can be used to create immersive experiences, including AR/VR, in video games and simulations. They also have use cases in healthcare, customer support and risk analytics. Breakthroughs include Google's Gemini Nano, Meta's MiniLM and TinyLlama.

09 RAM Shortage



Massive demand for AI hardware has led to a global crunch of memory chips (or RAM), with a small group of Asian suppliers controlling production. Memory chips like HBM and DRAM are critical for AI accelerators to function at scale.

Big tech is prioritizing supply even at high cost and this could make gadgets pricier. Google reportedly fired execs after they failed to bag supply agreements for HBM in South Korea.

10 Data Centres In Space

Data centres present challenges in the form of massive water consumption, carbon emissions and energy depletion. But in space, they can draw from round-the-clock solar energy and utilise the vacuum for efficient cooling.

This year, various tech CEOs spoke about making this bold bet a reality in the future. Google's Project Suncatcher, Nvidia-backed Starcloud and Elon Musk's SpaceX are all focused on this alternate path to sustainable computing.

11 Ethics & Guardrails

As Gen AI is being used widely, it has raised concerns of copyright infringement (with lawsuits against AI companies from musicians, authors and media companies), lack of transparency and accountability in data usage, cybersecurity risks, reinforcing bias, and tracking without consent.

Guardrails are needed to prevent misuse through data leaks and unauthorised action, ensuring compliance, and reinforcing human-in-the-loop approach for key decisions.