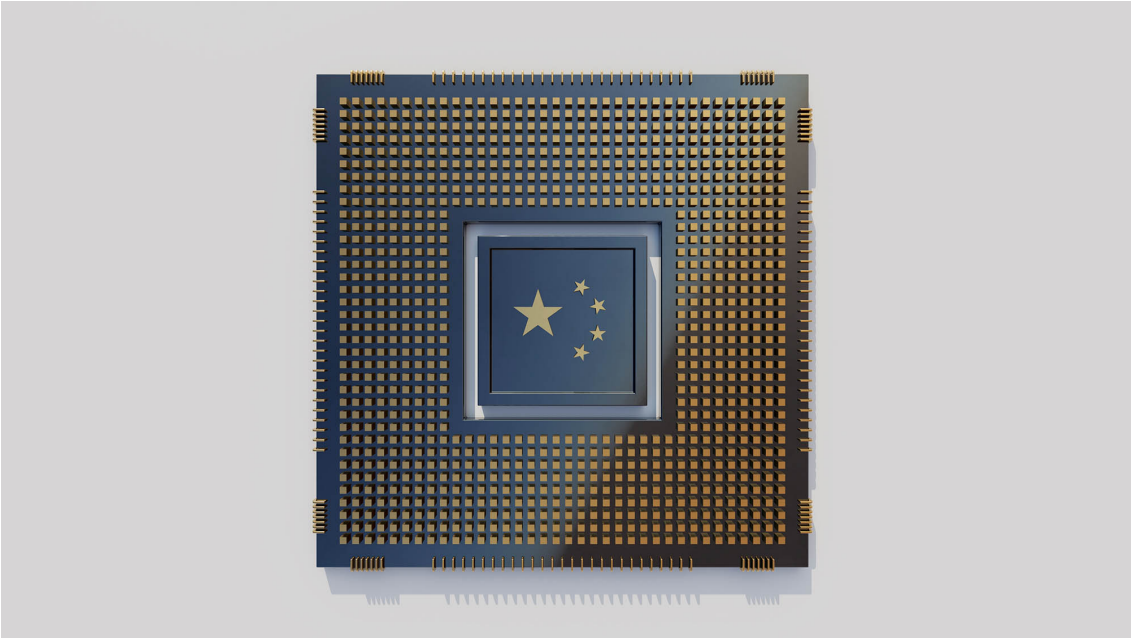
HARVARD BUSINESS REVIEW

Technology And Analytics

How Savvy Companies Are Using Chinese AI

by Amit Joshi, Mark J. Greeven, Sophie Liu and Kunjian Li

From the Magazine (September–October 2025)



Ricardo Tomás

OpenAI's November 2022 launch of ChatGPT caught Chinese technology companies completely off guard. Overnight, firms like Alibaba, Tencent, and Baidu, once competitive with global players like Google and Microsoft, had become laggards. But nearly three years later, Chinese companies have more than caught up to their U.S. rivals: They have forged a new parallel path with generative AI. DeepSeek, which was founded in 2023, [illegible] the AI space. In less than a year, and with a fraction of the computing

and data resources of U.S. models, DeepSeek-R1 is performing comparably to OpenAI's GPT-4o and Anthropic's Claude 3.5 Sonnet. Another startup, 01.AI, has launched the Yi-Lightning model, which rocketed up the leaderboard in terms of price, performance, and accuracy. But what's emerging in China isn't a clone of Western systems. It's a strategically distinct model of gen AI, adapted to thrive under different constraints and to meet different priorities. Firms such as DeepSeek are leveraging foundational advances while engineering distinct AI systems designed for cost-efficiency, rapid deployment, and targeted applications.

These developments are unfolding against the backdrop of mounting geopolitical pressures and export controls—particularly on semiconductor access—that were meant to cripple China's gen AI development. Instead, they catalyzed it. Today more than 300 gen AI services are registered with the Cyberspace Administration of China. Huawei, for instance, fast-tracked the development of its Ascend chip series, building a homegrown alternative to Nvidia's chips, and Ascend chips now power national-scale data centers.

Many global companies are integrating gen AI into their businesses, and most are using Western tools from companies like OpenAI, Google, and Anthropic. But they now find themselves navigating a second, very different ecosystem, one that has grown rapidly, quietly, and with a logic all its own. At the core of China's momentum is the country's commitment to building a modular and resilient AI infrastructure that enables rapid adaptation to local needs. Chinese firms are tackling challenges such as hallucinations, model costs (economic and environmental), and regulatory alignment with approaches that often diverge from those

resili I

chips to storage solutions, is attuned to local nuances. That approach differs sharply from the more generalized, broad-based research approach of the Western ecosystem.

The ongoing uncertainties surrounding U.S.–China tariffs further complicate the situation. Western technology, including large language models (LLMs) and infrastructure, is already hard to access in China. But now there is a possibility of restrictions on Chinese AI models in the United States, and the vetting of vendors will require more rigorous due diligence—from reviews of chip sourcing and cybersecurity audits to alignment with local regulations. Multinationals that operate in both regions must navigate a shifting landscape. Executives must weigh political risk, supply chain resilience, and governance compatibility alongside performance benchmarks when considering Chinese AI partners.

Western executives face a unique challenge. They can no longer assume that the best gen AI tools will all come from one ecosystem. Whether you're deploying customer service platforms, logistics optimization tools, vertical applications in healthcare or finance, or general-purpose AI agents, you'll most likely need to learn to work across Western and Chinese technology. Access to Chinese cloud providers, compliance with evolving AI governance, and compatibility with Chinese LLMs are fast becoming not just tech challenges but core business issues. Companies that fail to align their gen AI strategies with these new realities may find themselves outpaced by competitors who can move faster, for less money, and with better regulatory support. Navigating this terrain requires more than technical integration. It requires a strategic mindset.

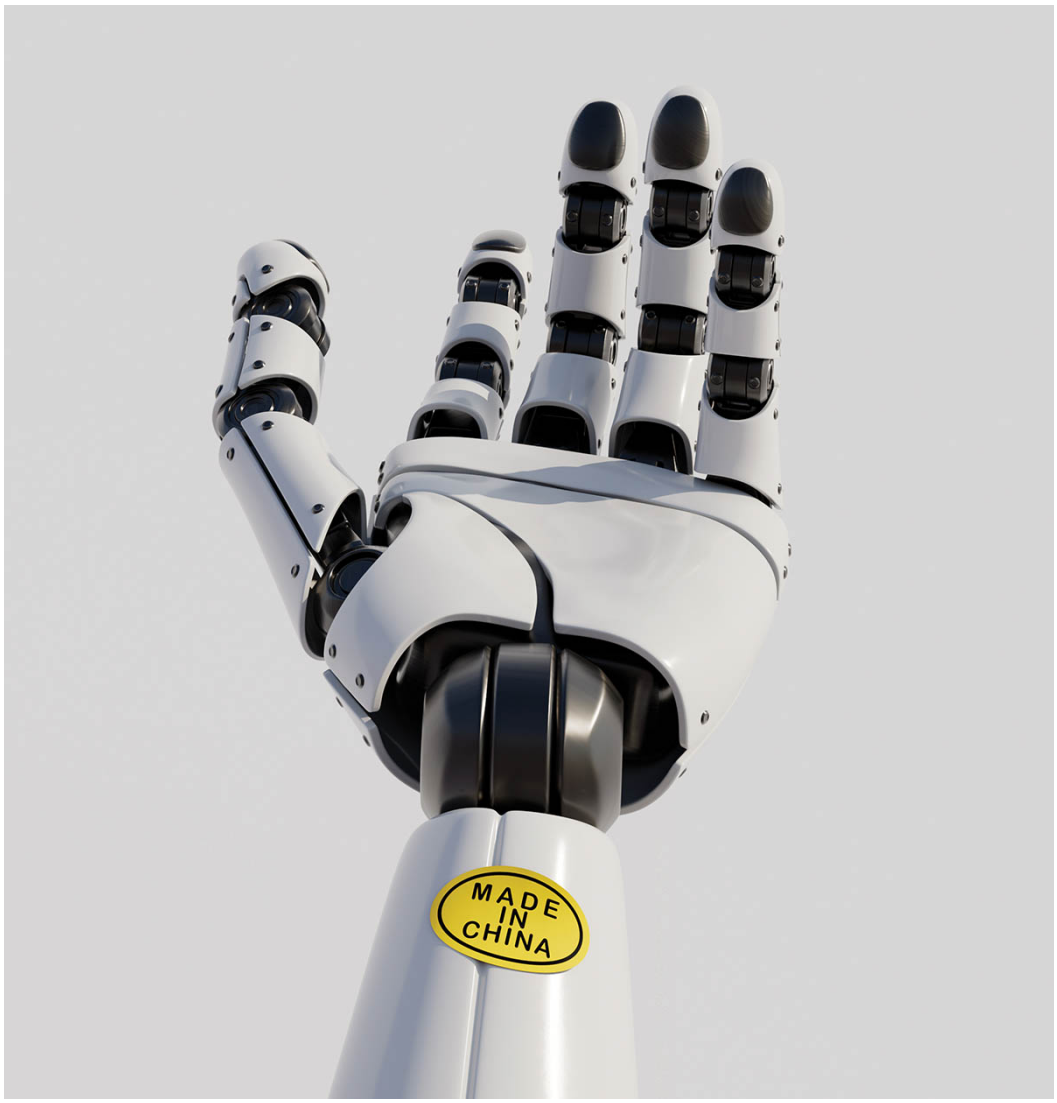
global firms understand China's unique gen AI architecture. As researchers and faculty members at IMD with extensive experience implementing AI strategy for organizations across multiple continents, we have worked closely with Chinese and Western firms to understand how gen AI is being developed and deployed on both sides of the globe. We have also conducted data-driven research on innovation and the future readiness of Chinese sectors such as apparel, pharmaceuticals, and technology.

In this article we present a road map for adopting a dual-track approach that leverages Chinese and Western gen AI to achieve superior strategic outcomes and operational efficiency. Global executives must keep three key issues in mind: First, China's ecosystem differs in terms of hardware supply chains and regulations. Second, there are real-world compliance constraints with Chinese technology. Finally, companies that fail to recognize how much more cost-efficient Chinese models are than most Western models will get left behind. This isn't about East versus West. It's about designing strategies that work in a world with more than one AI future.

The 3C Framework

Just as Amazon, Google, and Facebook did, China's tech giants—Alibaba, Baidu, Tencent, ByteDance, and Huawei—began in sectors like e-commerce, search, and social media but have since added world-class AI to their arsenals. Now midsize Chinese pioneers such as SenseTime and iFlytek are catching up in gen AI. In addition, a wave of fast-growing challengers—often referred to as China's "six little tigers" (01.AI, Zhipu AI, Moonshot AI, MiniMax, Baichuan AI, and StepFun)—is rapidly shaping the country's gen AI landscape. Meanwhile, others like DeepSeek and ModelBest are carving their own paths. (ModelBest has gained atten els.)

The gen AI stack has three layers: the infrastructure layer (storage and chips), the intelligence layer (LLMs), and the output layer (applications). However, Western and Chinese companies approach these layers differently. The Chinese gen AI players are leaning into *customization, cost leadership,* and *calibration* to build systems that maximize efficiency, prioritize real-world relevance, and embrace divergence.



Ricardo Tomás

In contrast to the decentralized ecosystem of U.S. AI tools, which often depend on multiple software vendors, Chinese AI companies have built integrated, end-to-end workflows. Consider Huaw
Minc

vertically integrated solution—comprising hardware, software, and optimization tools—can customize its infrastructure and adjust the models at a lower cost than it could using AI from a decentralized ecosystem. The implications are far-reaching. For AI providers, the rise of vertically integrated ecosystems highlights the strategic value of owning the full stack. For users, it underscores the need to choose wisely between open, decentralized architectures and closed, centralized alternatives. Business leaders must align their AI adoption strategies with their broader goals—whether that means maximizing flexibility and scale or prioritizing cost-efficiency, compliance, and system cohesion.

Let's look more closely at the first pillar, *customization in infrastructure.* Chinese companies have been innovating in this area for a few years now. We heard repeatedly from startup founders and senior executives that China's AI strategies are not about chasing general-purpose solutions. Instead, as we've noted, AI solution providers are building modular, adaptable infrastructure that is finely tuned to local technical, regulatory, and operational needs. Chinese subscription-based AI services empower small companies by delivering easy-to-install AI solutions. These solutions are engineered to optimize performance in local contexts, especially in the finance and healthcare sectors, where understanding fine nuances is crucial. Take Ant Group, which has built a set of AI doctor agents that are available through its Alipay app. The agents are powered by a healthcare-specific foundational model, developed in collaboration with leading hospital teams. What sets the app apart isn't just the data—it's the way the system has been trained to reflect how real doctors in China think and reason. By combining clinical literature, structured diagnostic data, and the decision-maki m that' ed.

That kind of contextual depth is hard for general-purpose models like ChatGPT to match.

Chinese companies believe that customized infrastructure is essential for achieving operational agility and responsiveness. Alibaba Cloud's storage solutions, for example, improve read and write speeds and offer greater flexibility, which is key for gen AI applications. By contrast, ChatGPT 3.5 relied on Microsoft's Azure cloud and used Nvidia chips—both technologies that predated gen AI and weren't optimized for it.

The second pillar, *cost leadership in model development,* also offers the Chinese advantages. In the West, immense resources have been poured into the development of gen AI models (estimates range from a few billion to tens of billions of dollars). But Chinese gen AI firms have built models with cost-efficiency as a design principle. While many Western companies pursue innovations and scale economies to drive efficiency, Chinese entrepreneurs believe that true cost efficiency isn't just about slashing expenses. It's about leveraging mature AI solutions rather than developing models from scratch, which is feasible in China's vast market with centralized AI solution suppliers. Western firms thus focus on building the most cutting-edge infrastructure and models with the understanding that they will eventually deliver business results. Chinese firms build for business outcomes; the models and infrastructure are a means to that end. This mindset has driven firms like Tencent and Baidu to adopt techniques that provide practical results with relatively little investment. China's vertically integrated platforms, which provide homegrown chips, cloud infrastructure, and model-as-a-service offerings, reduce training costs while supporting multilingual and multimodal applications. Of course, Chinese comp as

well as the failures) of Western competitors such as OpenAI, Google, and Meta.

> **This isn't about East versus West. It's about designing strategies that work in a world with more than one AI future.**

For companies facing budget and infrastructure constraints, this pillar can translate into real advantages: faster deployment, lower training and inference costs, and easier integration into localized, domain-specific applications. While Chinese firms have benefited from the fundamental models made in the West, their ability to adapt, refine, and deliver high-performing models at a fraction of the cost—especially in environments with limited access to high-end Western infrastructure—offers a compelling value proposition. Their constraint-driven ingenuity has fostered rapid, pragmatic innovation because cost discipline is not merely a strategic choice but also a survival imperative.

The third pillar, *calibration for real-world applications,* ensures that AI models are not only theoretically sound but also practically effective. Chinese firms are obsessed with making sure their models work—not just in theory but in real, dynamic environments, such as retail, finance, hospitals, and government offices. Chinese AI companies are constantly testing and iterating. Notable innovations include Moonshot AI's Kimi, which in March 2024 became the first AI model to process up to 2 million Chinese characters in a single conversation. This advance positioned Kimi at the forefront of AI in handling extensive amou
capa

practical, document-heavy use cases in sectors such as healthcare, education, and customer service. Moonshot AI is aligning technical ambition with strategic deployment.

The three pillars—customization, cost leadership, and calibration—set China's gen AI apart from the West's. Chinese innovators are not replicating Western models; they are building a purpose-driven ecosystem grounded in their own unique market needs and constraints. The 3C framework gives us a lens to understand these dynamics—not just as technical variations but as strategic alternatives. And it offers a way for global leaders to evaluate how they should engage with a fragmented but interconnected AI world.

What's emerging in China is a differentiated full-stack ecosystem—models, infrastructure, deployment platforms, and vertical applications—that's increasingly competitive on its own terms. For companies large and small, the strategic question is no longer whether to engage with China's AI ecosystem but how to do so effectively—whether through partnerships or pilots or by drawing lessons from a system built to prioritize speed, scale, and real-world utility.

The Future of Chinese Gen AI Applications

As Chinese gen AI rapidly caught up to its Western counterparts, its application ecosystem swung into action, quickly developing innovative applications and use cases. Gen AI applications can be broadly categorized into three main types: efficiency-enhancing tools, general-purpose software, and industry-specific vertical applications. Chinese companies currently excel in the third category, deploying AI across a wide range of use cases there, such as customer engagement and enterprise platforms. They're not just t ont lines of real businesses.

Some Chinese AI companies have created products with immediate practical applications, such as Baidu's Ernie Bot. Because its training incorporated structured knowledge, Ernie has demonstrated strong performance in tasks that require deep contextual understanding. For example, in customer service scenarios, Ernie can tap into company-specific knowledge graphs and regulatory frameworks to generate highly accurate, policy-compliant responses—a level of enterprise alignment that general-purpose models typically cannot deliver out of the box. Ernie is an example of how thoughtful model design, focused on real-world integration rather than theoretical scale, can produce high-performing, lower-cost models, which are especially useful for global companies managing tight AI budgets.

Ricardo T

Similarly, Alibaba has made its gen AI tool available to sellers on its platforms, including its e-commerce company Taobao. There are more than 200,000 suppliers on Alibaba's platforms, offering more than 200 million products. Alibaba's tool allows these suppliers to create high-quality, professional-looking portfolios for each of their products. A few years ago, smaller suppliers wouldn't have been able to make portfolios like these for any but a handful of the best-selling SKUs.

Trip.com, China's largest online travel platform, is pushing the boundaries of personalized travel through its proprietary LLM, Wendao. Trained on 20 billion high-quality travel data points and powered by real-time information and proprietary algorithms, Wendao enables the platform to tailor itineraries to individual preferences and adapt dynamically to changing travel conditions. AI now touches nearly every corner of Trip.com's business. Engineers are using AI to cut coding time by 15% to 30%, content creation time has dropped from 8.5 minutes to 15 seconds with a 98.9% quality pass rate, and marketing content is produced in just three minutes. On the service side, more than 60% of user inquiries are resolved through AI-driven self-service, reducing costs and enhancing responsiveness. All those improvements are part of Trip.com's strategy to drive growth through intelligent automation.

In logistics, SF Technology customizes foundational models for supply chain decision-making. Its Fengzhi model blends the conversational strength of LLMs with the precision of domain-specific small models, tackling common challenges like hallucinations and inaccuracy in decision-making. The result is a highly efficient intelligent agent designed for logistics operations. Its second model, Fengyu, is already deployed across more than 20 bu t, dispa 30%

of its training data to general-domain data and 20% to logistics-specific material (versus Western models that use 100% general content) in order to build smaller, domain-focused models that reduce inference costs while maintaining performance. It's a clear case of application-driven AI, optimized not for scale but for strategic depth in an industry.

In the healthcare sector, Medlinker's MedGPT, an AI doctor, has demonstrated professional-level diagnostic capabilities comparable with those of human physicians in a hospital setting. During patient visits, the model can quickly infer the likely causes of an illness from a patient's symptom descriptions. It provides preconsultation medical support and improves the efficiency of both inquiry and diagnosis. It also illustrates how rigorous real-world testing can drive tangible improvements in performance. Similar real-world calibration practices are used outside healthcare. Kingsoft Office WPS AI's smart assistant deploys AI in productivity applications, for example, and Xiaomi deploys it for customer service.

A Hybrid Approach

So far we've described the 3C framework that underpins China's gen AI ecosystem, but our research finds that a hybrid AI approach—a strategic integration of Chinese and Western AI solutions—is on the horizon. We believe we are entering an era when no single gen AI stack or tool will dominate.

The United States and China are now running side-by-side ecosystems with different strengths, constraints, and priorities: The Chinese system is producing high-impact models at scale and speed. The Western systems remain strong in frontier research and foundation model breadth. By taking a dual-track approach, organ

For example, Nestlé has been partnering with Chinese companies such as e-commerce giant JD.com, logistics innovator Cainiao, AI company iFlytek, and others to build localized AI solutions for consumer insight and product innovation. It has found Western LLMs to be more effective for applications in logistics and inventory management, but Chinese AI-powered platforms have provided higher-quality outcomes in customer analytics. Using a combination of AI technology maximized Nestlé's operational efficiency while preserving its competitive advantage in consumer insights. It also helped the company navigate complex regulatory landscapes and market dynamics.

> **Chinese innovators are not replicating Western models; they are building a purpose-driven ecosystem grounded in their unique market needs and constraints.**

Take Starbucks as another example. While Starbucks employs AI in both markets, the applications it uses are customized for and embedded into the two distinct ecosystems. In the United States, Starbucks focuses on enhancing personalization and operational efficiency through proprietary platforms like Deep Brew. In China, by contrast, Starbucks uses partnerships with tech giants such as Alibaba to facilitate the integration of AI into digital ordering and delivery services, in alignment with the country's mobile-centric consumer behavior. It has also established the Starbucks Innovation and Tech Center in Shenzhen to accelerate growth.

To na and both

about their own strategic direction. Additionally, they'll need to address questions about cybersecurity and data protection. We recommend that business leaders start by taking the following three steps.

1. Research both AI ecosystems. Keep a close eye on what's happening in China's gen AI ecosystem. Unique and innovative solutions have already emerged from this market, and more are sure to come. The challenge for Western business leaders is staying informed: Most non-Chinese media outlets cover only the most extraordinary developments in that market. Consequently, companies will need to set up systems to monitor technological, regulatory, and application developments in China.

Large organizations should consider exploring partnerships with Chinese AI companies to get an inside track on the latest developments and tools. The exploration process must include a deeper analysis of the data-sharing and governance practices of these models as well as the regulatory bases on which they operate.

You'll need to build an intelligence infrastructure for this step, one that monitors key Chinese and bilingual news sources such as Caixin, 36Kr, TechNode, and Rest of World. You should also work with specialized, feet-on-the-ground think tanks to help you analyze, interpret, and compare systems.

In addition, you'll need to benchmark Chinese models and use cases. Track major Chinese gen AI platforms, such as Baidu's Ernie Bot, Alibaba's Tongyi Qianwen, iFlytek's SparkDesk, Tencent Hunyuan, and Zhipu AI's GLM. Try to get firsthand experience in real-world applications in retail, manufacturing, educ

work with now organize market immersions in China for this purpose.

Finally, establish local partnerships and a tech scouting network. It may make sense to appoint a Chinese-speaking local innovation scout to monitor things and report back. This scout should be embedded in the local ecosystem—attending major events such as the Shanghai World AI Conference and the ZGC Forum in Beijing.

2. Evaluate Chinese systems. We can expect the Chinese ecosystem to produce new business models and monetization strategies—just as it did with mobile technologies, where Chinese innovations like social commerce and mobile wallets and payment systems surpassed equivalent developments in the West. The brutal competition and scale of this market allow for business models that may not be possible elsewhere.

Therefore, it is imperative for companies to study new business models closely and be ready to adapt those that are most readily transferable. (Of course, you will need an understanding of the foundational technology on which the business models run, since token costs, computing-power requirements, carbon footprints, and scaling would all be dependent on the underlying tech stack.) LVMH is an example of one multinational that has done this. It partnered with Alibaba to leverage its generative AI capabilities—including the Qwen LLM and Model Studio—to localize and optimize its digital retail operations in China. Understanding and adapting to the host country's AI infrastructure unlocked both performance and cost advantages for LVMH.

P

research strength and infrastructure of Western technology, leaders can unlock new avenues for innovation.

Procter & Gamble is an example of a company using AI in China to create hyperpersonalized marketing. P&G partnered with Douyin to leverage China's "interest-based e-commerce" model, which combines short videos, algorithmic discovery, and direct purchasing. Unlike Western e-commerce tools that focus on search and brand loyalty, Douyin leverages algorithmic recommendations, impulse buying, and entertainment-driven discovery.

P&G co-created products with consumers through live-streaming feedback loops and by testing new formulations. The insights gleaned enabled rapid adjustments to packaging, pricing, and messaging. The model relied on Douyin's AI infrastructure for analytics, dynamic pricing, and targeted delivery—and it required P&G to rethink product development timelines and localize data architectures and consumer touchpoints to align with Douyin's technology.

3. Combine Western and Chinese systems. Companies may end up running two parallel gen AI models and applications. It is likely that global companies may choose to run applications that require high levels of accuracy and transparency on more accessible and familiar (Western) models like ChatGPT, Gemini, or Llama. Those models may be better suited for applications in pharma, banking, and government, especially because these sectors demand high performance, have higher costs, and carry significant regulatory constraints and transparency requi

But for industries such as retail, consumer goods, and media, and for routine tasks like summarization, basic coding, and customer service, companies may prefer to use lower-cost models from China. Lower regulatory demands make them well suited to these use cases. For example, BMW plans to incorporate artificial intelligence from the Chinese startup DeepSeek into its new vehicle models in China later this year. This collaboration is part of BMW's strategy to strengthen partnerships for AI integration into vehicles; the company is enhancing features such as the Intelligent Personal Assistant to provide more-intuitive user experiences tailored to the Chinese market.

In April 2025, the German automotive supplier Bosch announced its first order for Chinese high-performance computers designed for AI-enabled vehicle cockpits. The system offers voice recognition, navigation, and driver assistance. Bosch plans to manufacture more than 100,000 units, demonstrating its commitment to meeting the high demand for AI-enabled vehicle technologies in China.

In the long run we may see a dual AI ecosystem emerge. But unlik ble outsi

ecosystems that will allow companies to pick the best of both worlds.

...

China has built a parallel gen AI stack—one that works differently, moves quickly, and serves a different set of needs. By combining the best of China's agile, deployment-ready models with the research strength and infrastructure of Western technology, companies can unlock new avenues for innovation and growth. This is not about choosing sides. It's about being ready for a multipolar AI future. Now is the time to assess, partner, and integrate.

A version of this article appeared in the September–October 2025 issue of *Harvard Business Review.*

Amit Joshi is a professor of AI, analytics, and marketing strategy at IMD. He advises global organizations on applying AI strategically and leads several executive programs on AI and digital strategy.

Mark J. Greeven is a Chinese-speaking professor of management innovation and the dean of Asia at IMD. He advises global organizations on digital ecosystems and AI strategy, drawing on two decades of research and collaboration with leading Chinese tech compa

Sophie Liu is a research associate at IMD in China.

Kunjian Li is a research trainee at IMD and a graduate student at the University of Chicago.





Read more on **Technology and analytics** or related topics **AI and machine learning**, **Generative AI**, **Information management**, **Strategy**, **Business management**, **International business** and **Asia**