Infosys
topaz

# AI IN CHINA: A CXO'S GUIDE TO NAVIGATE THE EVOLVING LANDSCAPE

**Abstract**

Artificial Intelligence (AI) is transforming industries across the globe, heralding a new era of enterprise potential. From the perspective of businesses, this change is impacting both AI providers and their consumers.

Globally, AI applications within various frameworks are in different stages of hype and deployment cycles. Deep learning, machine learning and computer vision technologies have already seen widespread enterprise and public consumption. Generative AI (GenAI) is also gathering steam for enterprises across the world.

Business leaders of global and local enterprises operating in China must recognize that the AI narrative, roadmap, and strategies differ due to the country's unique regulatory environment, tech ecosystem, and industry-specific demands.

Infosys®
Navigate your next

## The State of AI in China

In every CxO conversation in China, AI is at the top of the agenda—particularly for global Multinational Corporations (MNCs) benchmarking their performance against international standards and for China-based enterprises looking to demonstrate AI-driven productivity. The AI ecosystem in China is distinct, characterized by a rapidly evolving regulatory landscape and a robust network of providers and applications. China's unique local ecosystem adds complexity to AI applications.

In the U.S., larger number of AI applications are public facing. In contrast, China's AI applications are more focused on industry-specific scenarios.

A duality in AI ecosystems is emerging, driven by contrasting approaches between China and U.S./Western players. As generative AI use cases continue to evolve across industries, the divide is becoming more pronounced. While this duality adds complexity for foreign firms operating in China, many have already adapted to the reality of a decoupled digital world. With the U.S. and China dominating AI development, a new frontier of opportunities is opening up for enterprises operating in China— MNCs and China Private Enterprises (PoEs).

## Challenges and the Way Forward

*For businesses in China deploying AI, three primary challenges stand out:*

### Regulations: Navigating a Complex and Evolving Landscape

China has been proactive in establishing a governance and regulatory framework, beginning as early as 2017 with the "Next Generation Artificial Intelligence Development Plan," which outlined ambitions to become the global AI leader by 2030. This was followed by various government interventions, including notices, guidelines, and standards. The recent one has been the "Guideline for the construction of comprehensive system for national AI industry" published by Ministry of Industry and Information Technology (MIIT) as recent as July'2024.  The most significant development to date is the "Interim Measures for the Management of Generative AI Services," published in August 2023.

Additionally, specific laws cover data privacy, such as the Cyber Security Law (CSL), Data Security Law (DSL) and the Personal Information Protection Law (PIPL) and the Cross-border Data Transfer (CBDT) regulations. For AI providers, including developers of large language models (LLMs), these regulations require models to be registered and undergo compliance testing. The list of Compliant LLMs is published by China's central internet regulator Cyberspace Administration of China (CAC) and can be accessed on the CAC website.

These regulations present several challenges for AI providers:

- Increased Compliance and Regulatory Costs: Enterprises must make significant investments to ensure adherence to new regulations.

- Audit and Certification Requirements: These requirements may limit technology choices, potentially stifling creativity and innovation.

- Strict Data Security, Handling, and Processing Requirements: Solutions must meet stringent data security standards, especially for enterprises that rely on diverse and extensive datasets.

For businesses that consume AI services, these regulations help provide clarity on compliance needs. Authorized and registered models that have undergone Large Model Standard Compliance Assessment offer reliable options for deployment. Another aspect from enterprises deploying AI is that the regulatory environment is more restrictive on public usage,  and this constraint has led to a surge in internal AI deployments within the organizations for purposes such as optimizing workflows, enhancing productivity, and reinventing internal corporate functions.

## Availability of AI Service Providers and LLMs

China-based users face restricted access to many well-known global AI providers due to the "Great Firewall" and self-imposed restrictions by some providers, such as OpenAI. As a result, platforms like GPT-4, LLaMA (Facebook), Gemini (Google), and Claude are unavailable to China-based enterprises. Additionally, resources like Hugging Face are inaccessible.

On the other hand, several domestic Chinese models are available and widely used, creating a parallel ecosystem similar to the dual-stack structure seen in the broader internet economy between China and the US. For general-purpose models, options like Alibaba's Tongyi Qwen (110B parameters, 10M context length), Zhipu's ChatGLM (130B parameters, 128K context length), Yi-Large from AI, SenseNova from SenseTime to name a few are available for deployments. These models are continuously evaluated against global counterparts (e.g., MMMU benchmarks) and are rapidly improving.

China also has a distinct advantage in vertical, industry-specific models. Models like Antelope (manufacturing), Sensenova (medical and legal), and Lingyi from Baidu (life sciences) are now available for commercial use. Antelope is tailored for the manufacturing industry, supporting industrial documentation like reports and code, as well as value chain activities such as maintenance, fault diagnosis, material inspection, and statistical analysis. Similarly Jiushu (from JD) is targeted for online retail and offers capabilities like personalized recommendation, product sales forecasting, logs and click behavior analysis, content creation , SKU analysis through image matching. Such vertical models are gaining widespread adoption and usage.

For AI implementations, these localized models trained on local data are often the better choice. They offer relevant local context, are more cost-effective, and are deployed with implementation approaches like narrow transformers and retrieval-augmented generation (RAG), which provide more focused and superior outcomes.

Communities like AI-wise are emerging as local equivalents of Hugging Face, supporting research and development, but more progress is needed in this area.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

## Hardware and GPU Availability

One of the most discussed constraints for AI deployment in China is the lack of availability of high-performance AI-focused GPUs and restrictions on manufacturing and purchasing equipment to produce advanced GPUs. The unavailability of GPUs, such as NVIDIA's A100 and H100, limit the training capacity for LLMs.

These constraints have forced businesses to innovate and adapt. On the hardware side, Chinese manufacturers are heavily investing in R&D, with new GPUs on the horizon, such as Ascend 9x0 (Huawei), Hugon GPU (Higon), and YunFei (Innosilicon). NVIDIA has also developed China-specific GPUs to comply with local regulations. Meanwhile, architectural changes, such as scaling out lower-end GPUs, are being explored to meet LLM training needs.

Additionally, the development of smaller-scale, business-integrated models coupled with fine-tuning techniques like lower rank adaptation (LoRA) which reduces the number of trainable models, has picked up steam and acceptability. These verticalized models cater to specific verticals and use cases, reducing the need for hundreds of billions of parameters training overhead. By leveraging highly contextual, private core data—thanks to extensive industrial digitalization, IoT, and 5G deployments—enterprises can train these models efficiently without needing high-end GPUs.

Other important consideration is that the unavailability of high-end GPUs has a limited impact on AI inference compared to training outcomes. Also, from a cost perspective, the ongoing inference cost of AI implementation for China-based models are approximately 5- 15% of the global prevalent models.

## Solid Six: Strategic Recommendations for Enterprises

Here are few recommendations for China businesses as they go about mainstreaming AI:

1. **Leverage AI Arbitrage:** China-based MNCs can benefit from operating in both AI ecosystems, leveraging the best of both worlds through deployments across global datacenter, hyperscalers, and parametrized data transfer within the permissible limits of cross border data laws. This approach helps leverage best of capabilities of both the ecosystems.

2. **Adopt an Ecosystem-agnostic Architecture:** For AI practitioners and enterprise consumers, the architecture must remain agnostic to the underlying general-purpose model, allowing for a plug-and-play approach.

3. **Invest in Talent and Change Management:** Continuous investment in AI-ready talent is crucial. It's important to address both the development of talent for AI and the role of AI in enhancing talent, to alleviate any concerns. The focus should be on embedding AI initiatives throughout the enterprise.

4. **Readiness for Rollouts in APAC and Global South:** With global economic conditions driving a "move east" strategy for businesses, and given the investments made by Chinese hyperscalers in the Global South, there is a compelling case for evaluating and deploying the China AI ecosystem in countries beyond China.

5. **Data Readiness:** For mainstreaming of AI in businesses, readiness of enterprise data is key. This is one area that delays industrialization of AI .

6. **Responsible AI:** Responsible AI is a business imperative that integrates ethics into the core of operations, bolstering reputation and enhancing transparency. Upholding principles of trust and responsibility is essential, achieved through proper checks and safeguards across any AI ecosystem. The guiding principle of being responsible and secure by design helps mitigate unintended consequences for enterprises.

## Conclusion

China-based enterprises are uniquely positioned to capitalize and build on their experience and achievements in computer vision and deep learning. By leveraging digitalization, IoT gains, and treating data as a core asset, they can focus on industry-specific implementations to enhance productivity and drive innovation. MNCs, in particular, have the advantage of leveraging both ecosystems, and responsible AI deployment to realize the potential generative AI. While the challenges are real, solutions state-of-the-art are within reach. The key is to discern valuable insights from the noise, take decisive action, and begin the journey.

## About the Author

### Rajnish Sharma

Country Head - Infosys China

Rajnish is Vice President at Infosys and Country Head of Infosys China. He is also a Member of Board of Directors of Infosys China. He is currently based in Shanghai.

He is responsible for overall Business and Operations for Infosys in China, including Sales, Marketing and Service Delivery across multiple centers in China . With a passion for Cloud, AI & Automation led digital service offerings, he leads a team of professionals at Infosys that define and implement strategies for local and global multinational clients, to help them navigate their digital transformation journey. He is also responsible for talent management and scaling Infosys capability in the Greater China region. He is a team player, leads a diverse cross functional team and is a strong advocate of Diversity and Inclusion at workplace.

During his 23+ years with Infosys, he has worked alongside clients in multiple industry sectors, across the globe including in US, Europe, and Asia Pacific geographies. Rajnish has a Bachelor's degree in Computer Engineering from National Institute of Technology, Kurukshetra, India.

Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises and communities to create value. With 12,000+ AI use cases, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems.

**Infosys®**

Navigate your next

For more information, contact askus@infosys.com

Infosys.com | NYSE: INFY

Stay Connected