# Alibaba Cloud unveils major AI enhancements to support global developers



During its annual developer summit, Alibaba Cloud announced a series of significant enhancements to its artificial intelligence (AI) offerings, aimed at driving innovation across various sectors. Among these updates are new large language models (LLMs), advanced AI development tools, and upgraded cloud infrastructure, all designed to support global developers in their quests for growth.

Dongliang Guo, Vice President of International Business and Head of International Products and Solutions at Alibaba Cloud Intelligence, emphasised the company’s commitment to providing value to global developers. “Alibaba Cloud is committed to delivering real value to global developers through cutting-edge AI models, enhanced cloud infrastructure, and accessible support programs,” he stated. “Together, we aim to spark more AI-driven innovations, benefiting startups, enterprises, and industries altogether across the globe.”

A primary focus of Alibaba's initiative is the expansion of its proprietary Qwen LLM family. The newly released Qwen 2.5 series offers a range of parameter sizes from 7 billion to 72 billion and is available through APIs on Alibaba’s generative AI development platform, Model Studio. These models support various applications, including text generation, language comprehension, and translation. In addition to text-based models, Alibaba has introduced multimodal AI capabilities, exemplified by the Qwen-VL series, which focuses on image analysis, and Wanx2.1, also known locally as Tongyi Wanxiang, a visual generation model that converts text prompts into images. Moreover, Qwen-Audio has been revealed to specialise in audio language applications.

The company also launched Tongyi Lingma, an AI coding assistant powered by Qwen 2.5-coder. This tool enhances software development by offering features such as code completion and debugging assistance, which aim to streamline processes and boost productivity. Model Studio now features new AI development tools that improve system reliability and scalability for developers around the world. Key highlights include a workflow mechanism that breaks down complex tasks, a collaboration agent for planning and execution, and a retrieval-augmented generation (RAG) feature that improves model accuracy using external data sources. Furthermore, Batch Reasoning allows multiple prompts to be processed at once, improving efficiency, and AutoEval facilitates automated model assessment for optimal deployment. These tools are expected to be fully operational by the end of the month.

To support these advancements, Alibaba Cloud has also enhanced its infrastructure. The introduction of the 9th Generation Enterprise Elastic Compute Service (ECS) instances, set to debut in global markets in April, promises a 20% boost in computing efficiency compared to previous versions. The incorporation of elastic Remote Direct Memory Access (eRDMA) technology aims to improve performance for intensive workloads by up to 50%. Additionally, Alibaba plans to launch its Container Compute Service (ACS) internationally in January 2025, which integrates container technology with cloud resources, allowing developers to focus more on application development rather than infrastructure management.

Alongside these technological improvements, Alibaba Cloud has initiated the GenAI Empowerment Program, a targeted initiative for startups and developers using Qwen models. The programme offers free cloud credits to support development efforts, invitations to workshops and tech expos, and opportunities for collaborative marketing to enhance visibility in global ecosystems. This initiative aims to act as a launchpad for generative AI projects, connecting participants with a network of innovators while reducing barriers to accessing advanced AI tools.

International organisations are already adopting Qwen-based solutions to accelerate their operations in various sectors. For example, Japanese AI firm Axcxept has customised the Qwen 2.5 LLM to develop EZO, a lightweight AI model adept at coding, reasoning, and complex writing in Japanese. The CEO of Axcxept, Kazuya Hodatsu, noted, “Qwen 2.5 has significantly enhanced its ability to process Japanese, giving it a competitive edge over other models. With Axcxept’s proprietary training process, we have developed a Japanese LLM that delivers unmatched accuracy.”

OxValue.AI, a deep-tech spinoff from the University of Oxford, also utilises Qwen-based multimodal AI models to provide corporate valuation services. By analysing diverse data sources related to a company’s research and development and financial operations, OxValue delivers precise assessments. Professor Xiaolan Fu, founder of OxValue.AI, explained the value of these technologies, stating, “Processing diverse data sources is essential for our valuation services. With the support of Alibaba Cloud’s AI models, we’ve significantly improved the quality and efficiency of this process. By collaborating with Alibaba Cloud, we’re able to deliver greater value to our corporate clients.”

As competition in the AI sector escalates, Alibaba Cloud is positioning itself as a significant player by offering robust solutions that support developers, startups, and enterprises. By focusing on reducing technical and financial barriers, Alibaba aims to foster innovation across multiple industries, potentially altering the landscape of AI development globally.

Source: [Noah Wire Services](https://www.noahwire.com)

## References

* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Corroborates the announcement of new AI models, including Qwen 2.5 and the text-to-video AI model, as well as the revamped cloud infrastructure.
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Supports the details about the Qwen-VL series, Wanx2.1 (Tongyi Wanxiang), and Qwen-Audio models.
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Provides information on the AI coding assistant Tongyi Lingma and the new AI development tools in Model Studio.
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Details the enhancements to the cloud infrastructure, including the 9th Generation Enterprise Elastic Compute Service (ECS) instances and Container Compute Service (ACS).
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Explains the GenAI Empowerment Program and its benefits for startups and developers.
* <https://www.alizila.com/alibaba-cloud-enhances-global-partner-ecosystem-to-drive-ai-adoption/> - Supports the initiative to enhance the global partner ecosystem to drive AI adoption, including the Alibaba Cloud Partner Rainforest Plan.
* <https://www.alizila.com/alibaba-cloud-enhances-global-partner-ecosystem-to-drive-ai-adoption/> - Corroborates the collaboration with AI tech partners and channel partners to support global customers’ digital transformation.
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Quotes from Eddie Wu, Chairman and Chief Executive Officer of Alibaba Cloud Intelligence, on investing in AI technology and infrastructure.
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Details the performance and capabilities of the Qwen2.5-Max model and Qwen-VL model.
* <https://www.alibabacloud.com/blog/alibaba-cloud-unveils-new-ai-models-and-revamped-infrastructure-for-ai-computing_601622> - Explains the next-generation data center architecture, CUBE DC 5.0, and other infrastructure updates.