

Qwen2.5-VL, an advanced visual-language model, can analyze images, graphics, and videos lasting more than an hour. It can also pinpoint specific moments in videos and convert scanned documents into structured data. The model comes in various sizes, and the most powerful version, Qwen2.5-VL-72B-Instruct, is available on Qwen Chat, Hugging Face, and Model Scope.

Meanwhile, Qwen2.5-1M specializes in working with long documents, processing up to 1 million tokens — much more than typical AI models. This makes it ideal for summarizing and analyzing lengthy texts such as scientific papers or reports. Alibaba also released an optimized framework on GitHub to help developers deploy the model faster and at lower costs.