

Qwen2.5-VL, продвинутая визуально-языковая модель, может анализировать изображения, графики и видео длительностью более часа. Она также может определять точные моменты в видео и превращать отсканированные документы в структурированные данные. Модель выпускается в разных размерах, а самая мощная версия, Qwen2.5-VL-72B-Instruct, доступна в Qwen Chat, Hugging Face и на Model Scope.

Между тем Qwen2.5-1M специализируется на работе с длинными документами, обрабатывая до 1 миллиона лексем — намного больше, чем типичные модели ИИ. Это делает его идеальным для обобщения и анализа длинных текстов, таких как научные работы или отчеты. Alibaba также выпустила оптимизированный фреймворк на GitHub, чтобы помочь разработчикам запускать модель быстрее и с меньшими затратами.