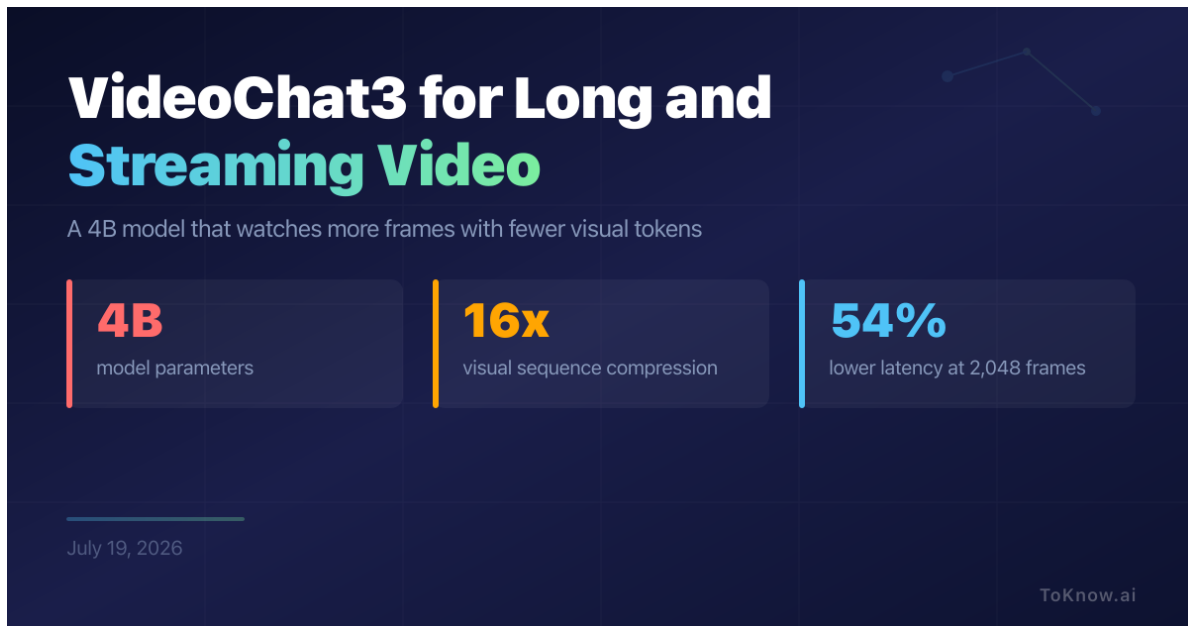


VideoChat3: 54% Lower Latency for 2,048-Frame Video

Kabui, Charles

2026-07-19

[Read at ToKnow.ai](#)



[VideoChat3](#) is a released 4B video-language model built to handle recorded and live video with one checkpoint. Its I3D-ViT encoder groups four frames and merges 2 x 2 spatial patches, cutting the visual sequence by 16x before it reaches the language model. In the authors' H200 test with 2,048 frames, that reduced latency from 44.45 to 20.41 seconds and memory from 106.92 to 80.78 GB versus Qwen3-VL-4B. The model also scored 72.3 on ODVBench, compared

with 59.9 for StreamForest, and 40.2 on VUE-TR v2, compared with 19.6 for Qwen3-VL-4B. These are author-reported results across different benchmark protocols, not an independent rerun.

For streaming video, VideoChat3 uses three states: Silence, Standby, and Response. Standby tells the next window to use more pixels because useful evidence may be arriving; Response returns to the cheaper setting after the model answers. This matters for an assistant watching a cooking demo, industrial camera, or accessibility feed, where processing every frame at full resolution wastes memory and answering early can be wrong. The team released [Apache 2.0 weights](#) plus 2.27 million academic, 116,200 long-video, and 617,183 streaming instruction records.

The release is not yet fully open despite the paper’s title. The [GitHub repository](#) currently contains the project website and inference demos, but its checklist still marks training code as unreleased and the repository has no licence file. The [data cards](#) also point users to original source datasets, whose terms must be checked separately.

Read More: [Video-MME-v2 Tests Where Video Models Still Fall Short](#)

Sources:

- [VideoChat3 paper](#)
- [VideoChat3 project page](#)
- [VideoChat3 model card and weights](#)
- [VideoChat3 code repository](#)
- [VideoChat3 models and datasets](#)

Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)