

Kandinsky 6.0 Video: Foundation Models for Synchronized Video and Audio Generation

Original summary — NOT the original paper or model card.

paper identified preprint · 2026-10-04 [7]

Authors/source: Team Kandinsky, Julia Agafonova, Bulat Akhmatov, Mikhail Aksyutin, Grigorii Alekseenko, Anastasia Aliaskina et al. (full author list in manifest)

Coverage: Video topic

What the source reports

A 3B Lite and 29B Pro diffusion-model family producing synchronized five-second audio/video. Dual-stream CrossDiT connects pretrained video and audio streams; super-resolution raises SD outputs to Full HD. [7]

Managing expectations

Five-second duration, native SD resolution, and remaining visual/audio-quality gaps against leading proprietary systems. Full HD is upscaled rather than native. The same report covers the super-resolution model. [7]

Original document and sharing status

An unmodified original paper is included separately under CC BY 4.0. This brief does not replace it.

Recorded rights: CC BY 4.0 [9]

Sources and downloads

[Primary source / manuscript details \[7\]](#)

[Official original document / PDF download](#)

[Hosted, unmodified original PDF](#)

[Complete document library](#) · [Read the research roundup](#)

Source briefs are Managing Expectations editorial summaries, not independent replications, peer review, or medical advice. Source claims and release details may change.