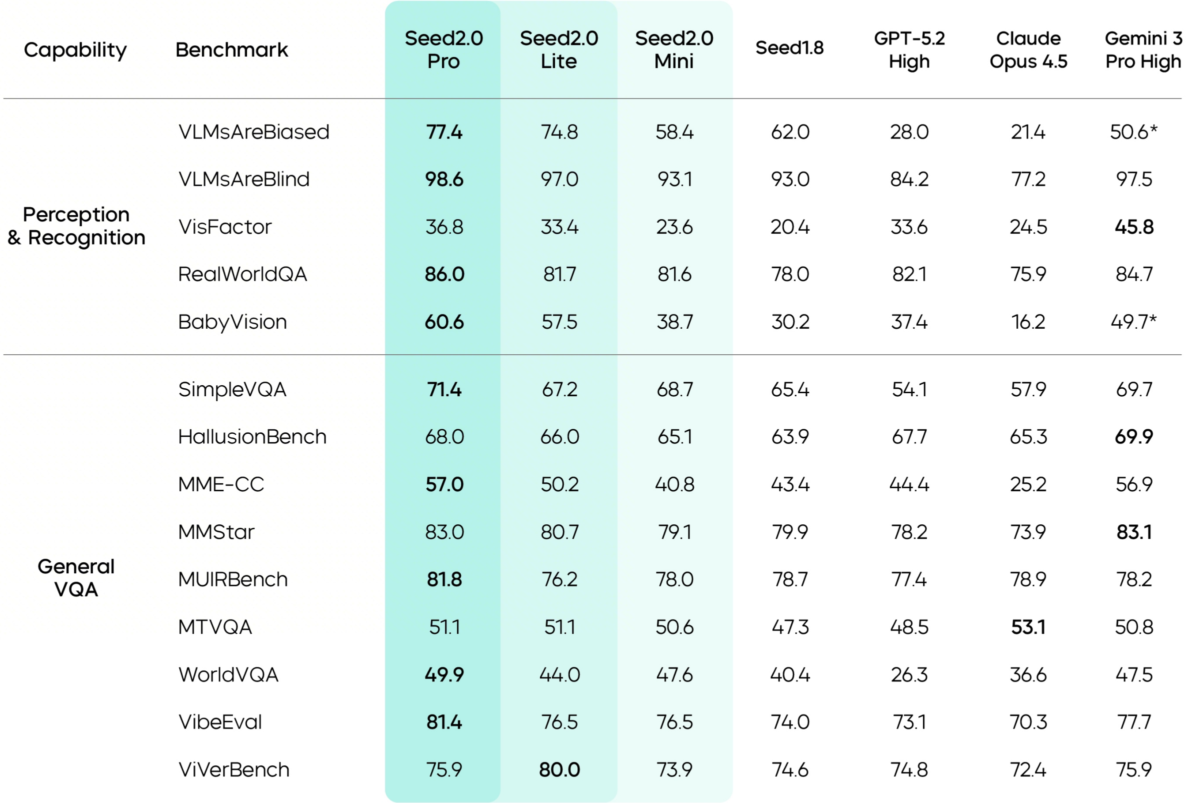
ByteDance Seedance: Multimodal Video Generation Reaches a New Threshold

Kabui, Charles

2026-02-16

Read at ToKnow.ai



Capability	Benchmark	Seed2.0 Pro	Seed2.0 Lite	Seed2.0 Mini	Seed1.8	GPT-5.2 High	Claude Opus 4.5	Gemini 3 Pro High
Perception & Recognition	VLMsAreBiased	**77.4**	74.8	58.4	62.0	28.0	21.4	50.6*
	VLMsAreBlind	**98.6**	97.0	93.1	93.0	84.2	77.2	97.5
	VisFactor	36.8	33.4	23.6	20.4	33.6	24.5	**45.8**
	RealWorldQA	**86.0**	81.7	81.6	78.0	82.1	75.9	84.7
	BabyVision	**60.6**	57.5	38.7	30.2	37.4	16.2	49.7*
General VQA	SimpleVQA	**71.4**	67.2	68.7	65.4	54.1	57.9	69.7
	HallusionBench	68.0	66.0	65.1	63.9	67.7	65.3	**69.9**
	MME-CC	**57.0**	50.2	40.8	43.4	44.4	25.2	56.9
	MMStar	83.0	80.7	79.1	79.9	78.2	73.9	**83.1**
	MUIRBench	**81.8**	76.2	78.0	78.7	77.4	78.9	78.2
	MTVQA	51.1	51.1	50.6	47.3	48.5	**53.1**	50.8
	WorldVQA	**49.9**	44.0	47.6	40.4	26.3	36.6	47.5
	VibeEval	**81.4**	76.5	76.5	74.0	73.1	70.3	77.7
	ViVerBench	75.9	**80.0**	73.9	74.6	74.8	72.4	75.9

Figure 1: ByteDance Seed Models

ByteDance's Seedance model family has evolved rapidly. Seedance 1.0, released in mid-2025, introduced a video foundation model built on a Diffusion Transformer architecture with multi-source data curation, video-specific RLHF (reinforcement learning from human feedback), and multi-stage distillation that achieved roughly 10x inference speedup. It generates 5-second 1080p video in about 41 seconds on an NVIDIA L20. Version 1.5 pro added a dual-branch Diffusion Transformer with a cross-modal joint module, enabling native audio-video generation with multilingual lip-sync and cinematic camera control. Seedance 2.0, the latest release, takes a unified multimodal approach: it accepts up to 9 images, 3 videos, and 3 audio files as combined input, producing 4 to 15-second clips with auto-generated sound effects. Its standout feature is reference-based control, where users can feed a reference video and the model will replicate its camera movements, lighting, and pacing while swapping characters or extending scenes.

The practical impact is significant. Seedance 2.0 compresses what previously required separate tools for video generation, audio synthesis, and editing into a single model. For advertising, short-form content, and storyboarding, this reduces both cost and turnaround time substantially. The model's ability to faithfully reproduce visual styles from reference material is so precise that Disney, SAG-AFTRA, and the Motion Picture Association have already issued cease-and-desist letters over unauthorized replication of copyrighted characters.

The copyright backlash highlights a deeper shift. Generative video models are now good enough that legal and ethical frameworks, not technical limitations, are the binding constraint on what gets produced. The gap between "technically possible" and "legally permissible" in AI-generated media has never been wider.

Sources:

- ByteDance Seed Models
- Seedance 2.0 Model Page
- Seedance 1.0 Technical Report (arXiv:2506.09113)
- Seedance 1.5 pro Technical Report (arXiv:2512.13507)
- The Decoder: Seedance 2.0 Coverage

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