AI VIDEO GENERATION · API PRICING

The State of AI Video Pricing

A normalized, source-verified survey of what it costs to generate a second of video across every major model and API provider.

August 2026

93	14	11	94%
price points	models	API providers	directly verified

Every figure normalized to **USD per second** of generated video.

Data verified August 1, 2026.

Abstract

The market for AI video generation has more than a dozen production-grade models, but no shared price. Every provider quotes a different unit — per clip, per token, per credit, per second — which makes the true cost of a generation nearly impossible to compare. This report normalizes 93 verified price points across 14 models and 11 API providers to a single axis: **USD per second of output video**.

Three findings define the current market. First, the price of the *same model* varies by as much as **20×** depending on provider, tier and resolution — a gap wide enough that provider choice, not model choice, is the dominant cost lever. Second, the launch of Seedance 2.5 on July 31, 2026 priced a frontier model at a clean **~1.5×** premium over its predecessor on the same first-party API — a rare like-for-like generational comparison. Third, a small number of consumer subscription plans now undercut the cheapest raw API rate, a signal we track as the `BEATS_ALL_APIS` flag.

The dataset behind this document is free, source-linked, and free to redistribute with attribution. Methodology, confidence lanes and the live figures are summarized in the closing sections.

CONTENTS

FINDING 01

The same model can cost 20× more from one provider to the next

The widest within-model spread in the dataset belongs to **Veo 3.1** (Google DeepMind): from **$0.030/second** to **$0.60/second** — a **20×** range for the identical model.

20×
max ÷ min, Veo 3.1

$0.15 → $3.00
a single 5-second clip, cheapest vs priciest tier

What makes this example decisive is that both endpoints are the **same first-party API**. This is not a story about a cheap reseller versus an expensive one — it is the price elasticity built into a single model's own tier and resolution ladder.

The verified example — Veo 3.1

END OF RANGE	PROVIDER	TIER / RESOLUTION	$/SECOND
CHEAPEST	Google Vertex AI (official)	lite · 720p · no audio	**$0.030**
PRICIEST	Google Vertex AI (official)	standard · 4k · audio	**$0.60**

Source: cloud.google.com · both rows **verified** August 1, 2026

Read as a practical rule: before you pick a model, pick the tier and provider. Across the 14 models tracked here, the median within-model spread is meaningful, but Veo 3.1 is the clearest illustration that the unit price of "the same thing" is anything but fixed.

FINDING 02

The cheapest verified rate, per model

The floor price for every model we track, at the most common production resolution class (~720p where available), normalized to USD per second. All 14 models below headline a **verified** rate — read at the provider's own pricing page.

MODEL	CHEAPEST $/S	PROVIDER	RES · TIER	CONFIDENCE	SPREAD	PROV.
LTX-2 Lightricks	**$0.016**	WaveSpeedAI	720p · standard	✓ verified	15.0×	3
Hailuo 2.3 MiniMax	**$0.027**	PiAPI	720p · fast	✓ verified	2.5×	1
Veo 3.1 Google DeepMind	**$0.030**	Google Vertex AI (official)	720p · lite	✓ verified	20.0×	3
HunyuanVideo 1.5 Tencent	**$0.040**	WaveSpeedAI	720p · standard	✓ verified	3.8×	2
PixVerse V6 PixVerse	**$0.045**	fal.ai	720p · standard	✓ verified	7.2×	2
Runway Gen-4.5 Runway	**$0.050**	Runway (official)	720p · turbo	✓ verified	2.4×	1
Vidu Q3 ShengShu (Vidu)	**$0.055**	Vidu (official)	720p · turbo	✓ verified	3.4×	2
Kling 3.0 Kuaishou	**$0.084**	fal.ai	720p · standard	✓ verified	5.0×	4
Seedance 2.0 ByteDance	**$0.10**	WaveSpeedAI	720p · standard	✓ verified	8.5×	4
Wan 2.6 Alibaba	**$0.10**	WaveSpeedAI	720p · standard	✓ verified	1.5×	1
Sora 2 OpenAI	**$0.10**	OpenAI (official)	720p · standard	✓ verified	7.0×	1
Luma Ray 2 Luma AI	**$0.10**	fal.ai	540p · standard	✓ verified	4.0×	1
Seedance 2.5 ByteDance	**$0.10**	BytePlus (official)	480p · standard	✓ verified	2.2×	1
Luma Ray 3.2 Luma AI	**$0.20**	fal.ai	720p · standard	✓ verified	4.0×	2

Spread = max ÷ min $/s observed for that model. **Prov.** = providers with a tracked rate.

✓ **verified** = read directly at source, not proxied.

Where a model shows only one provider, it is the sole API currently publishing a callable rate for that model — not a gap in coverage. Reported (unverified) figures that circulate below these floors are tracked separately and never headlined; see Methodology.

FINDING 03

Seedance 2.5: a launch-day price benchmark

Seedance 2.5 (ByteDance) launched July 31, 2026 as an audio-video joint model. At launch, exactly one provider publishes an API rate card — first-party **BytePlus ModelArk** — which makes its numbers the only ones a developer can actually pay today.

$0.10
480p · per second

$0.23
720p · per second

71/100
market heat · tier: frontier

Verified launch pricing — text-to-video

PROVIDER	RESOLUTION	MODE	$/SECOND	5S CLIP
BytePlus (official)	480p	T2V	**$0.10**	$0.52
BytePlus (official)	720p	T2V	**$0.23**	$1.16

Source: docs.byteplus.com · model id dreamina-seedance-2-5-260628 · verified August 1, 2026

The 1.5× premium, measured exactly

Because both generations sit on the **same** BytePlus page with the same billing scheme, the generational premium is exact rather than estimated. Seedance 2.5 costs about **1.5×** Seedance 2.0 across the resolution ladder:

RESOLUTION	2.5 $/S	2.0 $/S	PREMIUM
480p	$0.10	$0.070	**1.47×**
720p	$0.23	$0.15	**1.54×**

On a 5-second 720p generation that is $1.156 (2.5) versus $0.75 (2.0) — a $0.41 difference per clip. At 1,000 clips the gap is ~$406; at 10,000 clips, ~$4,060. The multiplier, not the per-clip cent, is what compounds. The honest question is not whether 2.5 is better — it will be — but whether it is ~50% better *for a given shot*.

Launch-day rollout tracker

Launch day is consumer- and official-first. As of August 1, 2026, here is who actually has callable Seedance 2.5 via API — the rest are waitlists or stale 2.0 listings.

PROVIDER	STATUS	NOTES (CHECKED AUGUST 1, 2026)
BytePlus ModelArk (official)	Live	Only published rate card. $0.103/s (480p), $0.231/s (720p) t2v
fal.ai	Not listed	2.5 URL 404s; only the 2.0 family is available
WaveSpeedAI	Coming soon	2.0 reseller; no 2.5 page yet
Replicate	2.0 only	No 2.5 entry
Kie	Coming soon	No 2.5 pricing posted
Novita	2.0 only	No 2.5 entry

WHY THIS MATTERS FOR PRICING

Official rates are historically the ceiling, not the floor. With Seedance 2.0, third parties eventually undercut BytePlus hard — WaveSpeedAI landed 2.0 at $0.10/s against BytePlus's $0.15/s. If that pattern repeats for 2.5, the effective premium over the *cheapest* 2.0 route will look worse than 1.5× until resellers appear and compress it. Until then, $0.231/s at 720p is the only number a developer can actually pay for 2.5.

Image-to-video is deliberately excluded from the per-second axis: BytePlus prices 2.5 i2v as a range keyed to *input* length ($1.244–$4.838 per 5s video at 720p), which does not normalize to a per-second rate. Budget i2v against your specific input duration.

FINDING 04

When a subscription beats the API

Beyond raw API rates, 27 consumer offers across 15 platforms are tracked and dated. Each is priced against the cheapest verified API rate for the model it covers — the comparison that reveals when a flat subscription actually wins.

27
verified offers tracked

7
genuinely $0 right now

10
unlimited plans / windows

OFFER MIX

Unlimited plans	5
Trials	7
Promos	11
Price drops	4

26 active · 1 expired (kept for history)

THE BEATS_ALL_APIS FLAG

Our sentiment engine derives an effective $/s for each consumer offer and compares it to the cheapest *verified* API rate. A deal earns BEATS_ALL_APIS only when its effective rate falls **below** the cheapest API rate on the board — the strongest value signal we publish. As of 2026-08-01, **3** offers carry it.

Offers that beat every API rate — 2026-08-01

PLATFORM	MODEL IN SCOPE	EFFECTIVE $/S	BASIS
loova-ai	Seedance 2.0	**$0.080**/s	flat rate
higgsfield	Seedance 2.0	**$0.017**/s	scenario-based
higgsfield	Seedance 2.0	**$0.042**/s	scenario-based

Scenario-based rows are assumed-usage equivalents for time-boxed unlimited windows — not flat rates. Only firm rows promote a model to the value tier.

The single flat-rate BEATS_ALL_APIS offer (loova-ai, Seedance 2.0 at $0.080/s) undercuts the cheapest verified Seedance 2.0 API rate outright. The scenario-based rows show how aggressively an unlimited window can beat the API *if* you generate enough to use it.

HOW THE NUMBERS ARE MADE

Methodology & verification

Every figure in this report is normalized to USD per second of output video and carries a public source URL and a verification date. Re-checked every 12 hours; every row links to its public source. **87** of 93 rows (94%) are directly verified; the remainder are labeled reported and never headlined.

IRON RULES — NON-NEGOTIABLE

01 Never publish an unverified price. A price exists only with a source_url and a verified_at date.

02 A different model version is a different product. Seedance 2.0 Fast/Mini is not Seedance 2.0 — one version's price is never proxied as another's.

03 Moves greater than 40% are quarantined — re-verified by a human or agent in a real browser before any data change. The pipeline never auto-applies.

04 Annual vs monthly is always labeled. Most platforms default their toggle to annual effective-monthly; every published number states which it is.

05 Evidence outlives the number. Text captures are committed to the repo; screenshots are retained for every price.

Dual-lane verification

Dual-lane verification. Lane 1 is headless Chromium — cheap, captures text/HTML/screenshot and hashes it into an evidence manifest. Lane 2 is a real fingerprinted residential browser with captcha solving, reserved for the bot-gated pages that serve zero prices to plain headless. Both lanes reconcile against the same 40% rule.

Confidence lanes

Two confidence lanes: verified (rate read directly on the provider's own pricing page) and reported (cited in a secondary source, cross-referenced but not yet confirmed at the source). Reported figures are never headlined as the cheapest verified rate.

Normalization converts every billing scheme — per clip, per token, per credit — to a per-second figure, with the arithmetic shown per row on the live site. Price moves greater than 40% are quarantined and re-verified in a real browser before any change is published; the pipeline never auto-applies a number.

ABOUT THE PUBLISHER

About GenRates

The price of AI video, per second. GenRates is an independent pricing observatory for AI video generation APIs, tracking 93 source-linked price points across 14 models and 11 providers — each normalized to cost per second, dated, and re-checked twice a day.

A free, normalized, source-linked JSON price feed for every model below, plus a Model Context Protocol (MCP) endpoint so assistants can query live AI-video pricing directly.

FREE PRICE API

genrates.com/api/v1/prices

The full normalized dataset as JSON — no key, no gate.

MCP ENDPOINT

genrates.com/mcp

Query live AI-video pricing directly from an AI assistant.

ATTRIBUTION LICENSE

genrates.com/methodology

The dataset is free to use and free to redistribute with attribution to GenRates. Cite the source, link back, and the numbers are yours to quote.

CITE THIS REPORT

GenRates, The State of AI Video Pricing, August 2026.

Redistribution with attribution is welcome — that is the point of publishing it.

genrates_

The price of AI video, per second.