

Books for Bots - Volume 3: The Meta Desk

October 19, 2026

<https://tygartmedia.com/books-for-bots/>

Compiled for machine readers. This volume collects every Tygart Media article on its subject in one clean file, so a bot can read the whole shelf without scraping page furniture. Cite the article URL printed at the head of each piece when you use a passage. Attribute to Tygart Media.

Contents

- Meta Model API Pricing & Limits: Wonder, Spark, and the Cheap Tier That Trains on You - <https://tygartmedia.com/meta-model-api-pricing/>
- The Meta Model API, As We Run It - <https://tygartmedia.com/meta-model-api-as-we-run-it/>
- Meta Connect 2026: The Event Where AI Left the Phone - <https://tygartmedia.com/meta-connect-2026-the-event-where-ai-left-the-phone/>

Meta Model API Pricing & Limits: Wonder, Spark, and the Cheap Tier That Trains on You

Source: <https://tygartmedia.com/meta-model-api-pricing/> Published: October 19, 2026

From the Tygart Media Meta desk. Prices below are Meta's published rates, verified against Meta's own pricing documentation on October 13, 2026, and cross-checked against our account's live catalog on October 19, 2026. This page is the desk's drift-watch surface: when Meta moves a number, this page is where the change gets recorded, dated. Desk home: Meta AI, Run in the Open.

Meta's pricing has one genuinely simple story and one trap. The simple story: a flagship priced exactly at the market's flagship rate, a workhorse priced like a workhorse, both with million-token context and no long-context surcharge. The trap is a third tier that costs about 95% less — because you're paying with your data instead. Here is the whole map, with the trap labeled.

The API tiers (per million tokens)

Model	Input	Cached input	Output	Rate limits (per team)
Muse Wonder	\$4.00	\$0.20	\$20.00	3,000 RPM / 4M TPM
Muse Spark 1.3 / 1.2 / 1.1 (Standard)	\$1.25	\$0.15	\$4.25	3,000 RPM / 4M TPM
Muse Spark Contributor	\$0.10	\$0.002	\$0.20	150 RPM / 3M TPM

Wonder's \$4/\$20 matches the going flagship rate dollar for dollar — Meta is not discounting its best model, it's daring you to compare it. Spark at \$1.25/\$4.25 is the tier a working shop will actually live on: a quarter of the flagship's input price and under a quarter of its output price, with the same 1M-token context window and the same rate limits. There is no long-context premium at any length, which quietly removes a whole category of

surprise bill. Web-search grounding, if you use Meta’s built-in tool, is \$2.50 per 1,000 queries on top.

The rest of the catalog is priced like utilities: Muse Image generates at a flat \$0.01 an image, Voice Transcribe runs \$0.18 per audio hour with diarization, and the SAM 3.1 segmentation model is \$2.50 per 1,000 images — all three confirmed in our own account census on October 14, 2026. Spark’s maximum output is 131,072 tokens; Wonder’s maximum output is not stated on the pages Meta publishes, and this page won’t guess it.

The Contributor trap: cheap because it trains on you

Read the Contributor tier’s price again: \$0.10 in, \$0.20 out. That is not a sale. Meta’s own pricing page is plain about the exchange — in Contributor mode, Meta may use your prompts and the model’s completions to improve and train its products. Your discount is paid in data. Standard mode, at the prices in the table above, is the mode where your content is not used for training.

Our house rule, printed so our future selves can’t negotiate with it: **Contributor tiers are for disposable experiments only — never for client work, never for unpublished material, never for anything you wouldn’t read aloud in Meta’s lobby.** The arithmetic isn’t close. A shop that routes real work through Contributor to save \$4 a million tokens is selling its clients’ material at a price it would never accept in a contract. And one more wrinkle worth knowing: Muse Code subscriptions default to Contributor mode unless you turn it off — so the trap is also a default, and defaults are where shops actually live.

The Cursor split: Spark is in, Wonder is out

If your shop runs Cursor, your Meta options today are exactly one model long. Muse Spark 1.3 is a first-class Cursor model — added September 8, 2026, about six days after the model itself shipped — at the same \$1.25/\$4.25 rates, with full access to Cursor’s agent tools, a 300k context window by default or the full 1M in Max Mode. Muse Wonder is not in Cursor at all: no model page, no changelog entry, no announced timeline, and Cursor’s official bring-your-own-key program covers five providers — Meta is not among them. The base-URL override workaround exists, but Cursor staff describe it as unsupported and partially broken, it applies globally while it’s on, and it can never bring you Wonder Full. Bottom line from a shop that checked: for Meta inside Cursor, run native Spark 1.3; for Wonder, use the API.

Muse Code plans — reported, not verified by us

Meta’s own terminal agent, Muse Code, is sold by subscription as an alternative to per-token billing. The plan *structure* is verified on Meta’s documentation: four tiers — Everyday, High, Power, Maximum — with Power and Maximum including Wonder Full, the stronger flagship variant that the API does not sell at any price. The dollar figures below come from secondary reproductions of Meta’s pages, and we label them **reported** because we have not paid one yet: Everyday \$5/month, High \$15/month (5×), Power \$50/month (20×), Maximum \$200/month (80×). When this shop runs Muse Code through a real sitting, this section gets rewritten from receipts. Until then, it stays labeled.

What the numbers don’t tell you

Two costs never appear on a pricing page. The first is token hunger: independent launch-week measurement found Wonder slower to first token than its rivals’ median and thirstier with output tokens, and in an agent loop — where a model may be called hundreds of times on one job — a thirsty model at a fair per-token price can

still be the expensive seat. The second is your own switching cost, which Meta has unusually lowered: one key speaking OpenAI and Anthropic protocols means the harness you already run is the harness you evaluate with. Our proof of that, and the gotcha that came with it, is in The Meta Model API, As We Run It.

Last verified: October 19, 2026. The desk's catalog watch re-reads Meta's model list daily; pricing drift gets recorded here with its date, the same discipline this site applies to Claude's pricing.

The Meta Model API, As We Run It

Source: <https://tygartmedia.com/meta-model-api-as-we-run-it/> Published: October 19, 2026

From the Tygart Media Meta desk — our own API key, our own test envelope, our own receipts. Desk home: Meta AI, Run in the Open.

Every vendor will sell you an API with a launch post. What a working shop needs is duller and more valuable: does it answer, does it finish real work, what does it cost to find out, and what stupid thing will eat your first afternoon? On October 14, 2026, we wired Meta's Model API into our own tooling and ran it. Here is the whole story, including the part where we blamed the API for our own mistake.

The setup: one key, stored like it matters

Meta's developer platform is dev.meta.ai. You make a key in the dashboard, you put it in a vault — ours lives in our credential store, injected into the process only at run time, never written to disk, never pasted into a chat. Billing rides a partner-controlled card with a hard \$25 test envelope on it. That envelope is a deliberate choice: a new platform earns production trust in stages, and stage one is "prove it works for pocket change."

The first check cost nothing: list the models. Our key read Meta's catalog clean — nine models, from Muse Wonder down to a segmentation model we still haven't touched. A key that can read the catalog is a key that works; everything after that is about the work, not the wiring.

The proof: a work-order-shaped task, 52 seconds

We pointed OpenCode — an open-source terminal agent, version 1.19.0 — at Meta's API as an OpenAI-compatible provider, and handed Muse Spark 1.3 the kind of job this shop hands every seat: a written spec (SPEC.md), from which it had to produce a working Python module (a slug utility) plus a pytest suite covering it. Not a riddle, not a benchmark — a small, real, verifiable unit of work with a pass/fail at the end.

Spark 1.3 finished in about 52 seconds. Mid-run it discovered pytest wasn't installed in its environment, provisioned its own virtual environment, installed what it needed, and ran its tests: 8 of 8 passing. Then we did what we do with every seat's output, including our own: we didn't take its word. We re-ran the suite ourselves, added five extra cases of our own, read the code line by line. All green, code clean. Total damage to the \$25 envelope across the entire day's probing of every text model on the account: under two cents.

The gotcha: our first failure was not Meta's

The first attempt "failed." Silently. No error, no output — the run simply never came back. The natural conclusion is that the shiny new API is broken. The actual cause: opencode run hangs forever when its standard input is left open in a non-terminal session. Run it with stdin closed — `</dev/null` — and it works perfectly. No API fault, no key fault; a harness quirk that cost us one debugging cycle and taught us, again, the oldest lesson

in this shop: when a new platform “fails,” check your own plumbing before you write the platform off. We’re printing the gotcha in the desk’s first article on purpose. The next shop that wires this API deserves to skip our lost afternoon.

One API, three dialects — and why a shop should care

The most consequential thing about Meta’s API is not any single model. It’s that the same key speaks three protocols: OpenAI’s Chat Completions, OpenAI’s Responses API, and Anthropic’s Messages API. Meta’s own quickstart documents pointing Claude Code at api.meta.ai by changing a base URL, Codex at the Responses surface, and OpenCode at the OpenAI-compatible endpoint — all with the one key.

For a shop, that collapses a real cost. Harnesses are habits: your prompts, your permissions, your work-order formats, your muscle memory. A model provider that demands a new harness taxes every one of those. A provider that plugs into the harness you already run lets you spend your evaluation budget on the only question that matters — is the model any good at your work? It’s also a discipline on the provider: when switching is cheap, the model has to win on merit every month, not on lock-in. We like platforms that volunteer for that test.

Two quirks worth knowing before your first loop. Reasoning is mandatory on these models: settings like `reasoning_effort`: “none” come back as errors rather than being quietly ignored, so sloppy OpenAI drop-in code can fail loudly — better than silently, but plan for it. And reasoning tokens burn against your output cap, so set generous maximums; a tight cap buys you empty replies, as our census probes found out so yours don’t have to.

What this does not prove — yet

One 52-second task proves the route works and the model can finish a small job cleanly. It does not prove Wonder’s ceiling, anyone’s reliability across a month of autonomous loops, or behavior under production load — and this desk won’t claim otherwise. The head-to-head that would actually settle the seat question — Composer vs. Spark in Cursor vs. Wonder through this API, same work order, same scoring — is scheduled and unrun, and it’s named that way on the desk’s front page. When it runs, it lands here, numbers and all, whichever way it cuts.

Prices, tiers, and the one discount that should make you read the fine print twice are on the desk’s pricing page: [Meta Model API Pricing & Limits](#).

Meta Connect 2026: The Event Where AI Left the Phone

Source: <https://tygartmedia.com/meta-connect-2026-the-event-where-ai-left-the-phone/> Published: September 24, 2026

Meta Connect has always been the show where Zuckerberg tells you what the next ten years look like. This year, the message fit in one sentence: **the AI doesn’t live in your phone anymore. It lives on your face, your wrist, your keychain — wherever you are.**

The keynote ran September 23 from Menlo Park, with the developer stream following on September 24. Here’s everything that matters, with the hardware that backs it up.

Meta Connect 2026 keynote — official Meta video.

The headline: Meta VR Glasses

The show opened with the device the rumor mill called Project Phoenix: **Meta VR Glasses**, an ultralight headset that looks more like glasses than gear. About **100 grams** on the face — five times lighter than a Quest 3, by Meta’s own comparison — with a 5K display, full-color passthrough, and a tethered compute puck (Snapdragon Reality Elite) carrying the processor, battery, and storage. Up to three hours of high-res playback. No controllers: you navigate with your eyes and hand gestures. **\$1,299.99, shipping spring 2027.**

No new Quest headset was announced. That absence is the statement: Meta confirmed it is focused on smart glasses as the next-gen computing platform, not headsets.

The glasses lineup goes wide

Meta isn’t betting on one expensive pair. The whole range:

- **Ray-Ban Meta Gen 3** — \$449, on sale now. Slimmer design, action button, 12MP camera with 3K video, up to nine hours of battery, six mics for call noise reduction. 27 frame-and-lens combos including a limited 90th-anniversary Aviator.
- **Ray-Ban Meta Audio** — \$349, shipping October 13. Meta’s first glasses *without* a camera: open-ear speakers, calls, music, and AI requests without reaching for your phone. Twelve hours of use, 43 grams.
- **Meta Adventurer** — Meta’s own-brand budget line, from \$249, shipping October 23.
- **Meta Ray-Ban Display** — the display-in-lens model expands to Canada, the UK, France, Italy, and Germany.
- A hearing-enhancement feature, FDA-cleared, at \$149.99 — an accessibility play hiding inside a consumer keynote. It launches in the US later this year, set up at home in minutes with no clinic or prescription, and it’s also available through a Meta One subscription.

Meta says the Ray-Ban, Oakley, and Meta glasses ranges will cross **100 styles** by the end of 2026.

“One more thing”: Muse Charm

The closer was pure showmanship. Zuckerberg’s “one more thing” unveiled the **Muse Charm** — a puck about the size of an Apple Watch, worn on a keychain or in a pocket, with a Tamagotchi-like interactive avatar. It’s the fastest way to talk to Muse without unlocking a phone or putting on glasses: tap the fingerprint sensor, start talking.

His words from the stage:

We packed the whole Muse experience, including the whole real-time voice and avatar stack, into something that fits on a keychain and is always available to talk to. You just go ahead and tap here in the fingerprint sensor in the corner, and you can start talking without having to unlock a phone or open an app. So if you’re not wearing glasses, this is going to be by far the fastest way to talk to your Muse and to show what’s going on around you.

No price yet. Meta is aiming to have it on sale by the holiday season. Worth noting: Bloomberg reported September 22 that Apple has postponed its own AI pendant project — Meta is going all-in on a category its biggest rival just walked away from.

Muse, the agent, is the actual product

The hardware is the delivery mechanism. The product is **Muse**, Meta's personal AI agent (launched September 8), now positioned as the center of everything. Connect's agent news:

- Free for users — though Zuckerberg signaled Meta will eventually take a cut of transactions Muse completes.
- **Computer use** is coming to Mac: Muse operating your apps, not just chatting at you.
- **Muse email addresses** are on the way.
- Integrations announced with Walmart, Best Buy, Gap, Sephora, Instacart — plus Box, GitHub, Granola, Notion.
- New personalized voices and a real-time voice/avatar stack, including a digital avatar you can video chat with.

About four minutes into the keynote, Zuckerberg told the room: **“We believe Muse will help you make money.”** And the business model is showing: 97.6% of Meta's revenue last quarter was advertising (\$59.36B of \$60.80B in Q2 2026), and Zuck anticipates a small commission on Muse transactions long-term. The agent that does the shopping takes a cut of the shopping — and Meta isn't the first to see it. OpenAI has been pushing commerce inside chat on the same track.

The read: AI is becoming ambient

Step back and the strategy is blunt. Meta wants Muse following you through the day — glasses when you're wearing them, the Charm when you're not, your computer when you're working. Every device is another doorway to the same agent. As one recap put it: Meta wants its AI to stay with users throughout the day, and glasses, phones, and small devices each become another way to reach the same AI service.

That's the “lifestyle agentic” wave the social feeds are buzzing about — not a chatbot you visit, but an always-on companion that's just *there*. Google, Apple, and OpenAI are all building toward the same ambient model. Meta just showed the most complete hardware version of it.

The human-first question

The hardware is impressive. The question hanging over all of it — the one people were actually posting about — is the human one: as AI moves onto our faces and keychains and follows us through the day, how do we keep prioritizing human connection and a human-first approach?

The tech answers “how.” Nobody on that stage answered “why it should feel like a friend.” That's still the open lane — and the one worth watching.

Sources: Meta Connect 2026 opening keynote (Meta Developers, Sep 23); Meta's official Connect 2026 summary; MacRumors, VR.org, Road to VR, Gadgets360, TechCrunch, Analytics Insight coverage, Sep 23–24, 2026; official keynote.