

# Toxic Virality: When Your Best Growth Asset Starts Billing You

Second Order Labs™ · June 2026

## ABSTRACT

Midjourney once shut off its own free trials to survive its own virality. That inversion explains why freemium is dying, and who gets to inherit it.

*Keywords:* Freemium, Generative AI, Usage-Based Pricing, Reverse Trial, Compute Costs, Product-Led Growth

*“Freemium is an acquisition model, not a revenue model. If your marginal costs aren't near zero, it will kill your business.”*

— Patrick Campbell, Lenny's Podcast

## I. When Virality Became a Liability

In March 2023, Midjourney did something no growth team would have pitched a year earlier: it shut off its own top of funnel. The company paused free trials because of what it called "extraordinary demand and trial abuse"<sup>[1]</sup>. The most viral consumer AI product of

user costs real money to serve, software's old growth assets turn into liabilities. Frictionless onboarding starts running up a tab, and GPU-heavy free users can rack up costs long before they show any intent to pay. Virality becomes a good way to lose money faster.

Freemium worked because one more user used to cost



We use cookies on this site to enhance your user experience. For a complete overview of all cookies used, please see your [personal settings](#).

Accept all

Decline

Customize

## II. Freemium Was a ZIRP Subsidy, Not a Law of Software

Freemium was never some natural law of software. It was an acquisition tactic that only worked under two conditions: near-zero marginal cost and cheap capital. For about a decade, both were available. Venture money subsidized user growth, and the cloud kept the cost of serving one more free user low enough to ignore. Once rates rose and compute got expensive, freemium stopped penciling out.

*"Freemium is an acquisition model, not a revenue model. If your marginal costs aren't near zero, it will kill your business."*

Patrick Campbell's line gets the trap exactly right.<sup>[2]</sup> Founders kept talking about freemium like it was a revenue model when it was really a funnel tactic. Kevin Hale made the same point in his Y Combinator

pricing lecture: don't use freemium unless your market is huge and the cost to serve free users is close to zero.<sup>[3]</sup> Those conditions were treated like defaults. They were never defaults.

Then the macro backdrop changed. Bessemer's State of the Cloud 2023 said the era of growth at all costs was over, replaced by a demand for efficient growth and profitability.<sup>[4]</sup> Gergely Orosz translated that into plain company math: once zero interest rate policy ended, subsidizing user growth with venture capital got a lot harder.<sup>[5]</sup> Heroku and PlanetScale were early proof that free tiers were already breaking before AI inference costs hit. Heroku killed free product plans in November 2022 and pointed to fraud and abuse.<sup>[6]</sup> PlanetScale shut down its Hobby tier so it could focus on paying customers and build a sustainable business.<sup>[7]</sup> AI did not invent this problem. It just made it impossible to ignore.

Figure 1: A free tier under high marginal cost is an open vault, not a funnel.

## III. Why Generative AI Breaks the Zero-Marginal-Cost Assumption

Generative AI breaks freemium because every query has a real compute cost, and that cost compounds with usage. Traditional software never had that problem at this scale. Serving the millionth user looked a lot like serving the first. Sam Altman said it plainly about ChatGPT: "we will have to monetize it somehow at some point; the compute costs are eye-watering."<sup>[8]</sup> The CEO of the biggest AI product in the market was warning that free had an expiration date.

a16z put AI gross margins at 50 to 60 percent, versus 60 to 80 percent or more for SaaS. That gap forces a repricing of free access.

*"In many cases, AI companies simply don't have the same economic construction as software businesses... Gross margins for AI companies are often in the 50-60% range, well below the 60-80%+ benchmark for SaaS." Martin Casado & Matt Bornstein, The New Business of AI (a16z)*

Figure 2: AI companies operate 20-30 points below the SaaS benchmark, dragging them toward services economics.

Lower margins push AI startups toward selling outcomes, not unlimited software access. Sequoia pointed to the core tension: the cost of creation and

knowledge work is falling, but the compute needed to deliver that work is not cheap.<sup>[9]</sup> Sarah Tavel took that one step further and argued that AI startups will

increasingly sell the work itself, not just access to a software tool.<sup>[10]</sup> Once that happens, a free tier means doing billable work for non-paying users. Ben Thompson's point is the one that matters most for

competition: unlike the web, AI answers carry real marginal cost, so scale can punish you unless every unit of usage is paid for.<sup>[11]</sup>

Figure 3: The shift from SaaS to AI flips the marginal cost assumption that made free tenable.

#### IV. Intent Underwriting: Growth Teams Become Risk Underwriters

If a free user can burn through dollars of inference before showing any buying intent, the funnel is no longer a funnel. It is credit risk. Under **Intent Underwriting**, growth teams treat each free signup as a compute loan. The question shifts from "how do we remove friction?" to "how likely is this user to pay back the spend we're about to incur?" A waitlist is not just scarcity theater. It is rationing for expensive compute.

Midjourney's free-trial pause looks a lot more rational through that lens. In a zero-marginal-cost product, a viral spike mostly creates storage and support headaches. In a high-compute product, a spike of bots and low-intent users can look like a default event. Abuse drives both GPU spend and support load.

Baremetrics described the support side of that problem when it explained why it dropped its free plan: "Free plans attract a lot of noise, support burden, and very little revenue."<sup>[12]</sup> Add GPU bills to that same traffic, and "noise" becomes a direct margin hit.

*The free signup is no longer a lead. It is an unsecured loan of compute, and most of those loans never get repaid.*

That changes the metric that matters. Raw signup volume stops being impressive. Intent-weighted signups matter more, because they tell you how much incoming traffic is actually worth serving.

#### V. The Models Replacing Freemium: Reverse Trials, Usage Pricing, and BYOK

The market is settling on three ways to cap compute exposure: reverse trials, usage pricing, and BYOK.

Model	What it caps	Psychological lever
Reverse trial	Time-boxes premium access, then drops to a thin free or paid floor	Urgency of a trial plus retention of freemium
Usage-based pricing	Aligns revenue directly with variable compute cost	Pay for what you consume
BYOK (bring your own key)	Pushes the inference bill onto the user's own API key	"Free" product, user-funded compute

Reverse trials work because they create urgency without leaving an unlimited free tier open. Elena Verna describes them as giving users "the best of both worlds: the urgency of a free trial and the long-term

retention of freemium."<sup>[13]</sup> Kyle Poyar's point is the broader one: freemium gets much harder to defend when capital is expensive and infrastructure costs are rising.<sup>[14]</sup>

Usage pricing solves a different problem. It protects you from the power user who can torch your margins under a flat seat price. OpenView has been direct about where the market is heading and treats usage-based pricing as the new default, not some passing fashion.<sup>[15]</sup> Seat-based pricing breaks in AI for a simple reason: one heavy user on a single seat can generate enough compute cost to wreck the old cost-per-seat model.

BYOK goes even further. It keeps the product free at the surface while pushing the inference bill onto the user's own API key. Tomasz Tunguz made the underlying economics clear: generative AI companies will carry lower gross margins than SaaS because querying LLMs costs real money.<sup>[16]</sup> If the startup cannot afford to eat that cost, the user will.

## **VI. The Second-Order Effect: Freemium Becomes a Big Tech Weapon**

The biggest effect is not that free tiers are disappearing. It is that startups are losing the right to offer them. Once free access gets too expensive for a startup to fund, only companies with giant balance sheets can keep doing it. Free stops being a startup growth move and turns into a moat for incumbents. Microsoft, Google, and Meta can run free AI features

as loss leaders because the inference bill barely dents the broader business. For a startup, that same bill can kill the quarter.

The people who get squeezed first are easy to name. Students, hobbyist developers, and open-source maintainers were the quiet winners of the old zero-marginal-cost model. They convert badly, which makes them the first group cut when efficient growth becomes the rule. That removes cheap access from the edge of the market, which is usually where experimentation starts.

Companies still have to manage the transition well. The cautionary story here is about communication. Docker's attempt to end its Free Team organizations produced such a backlash that the company had to publish a post titled "We apologize: We did a terrible job announcing the end of Docker Free Teams."<sup>[18]</sup> GitLab raised Premium pricing and got through it.<sup>[19]</sup>

Docker showed that ending free access is survivable. Botching the announcement is what triggers the revolt.

Within three years, the default AI signup flow will look less like old SaaS onboarding and more like credit screening. Treat free signups as costless wins, and the GPU invoice will correct the delusion fast.

## KEY FINDINGS

Midjourney paused free trials in March 2023, citing extraordinary demand and trial abuse. The most viral AI product had come to fear its own virality.

AI companies post gross margins of 50-60%, well under the 60-80%+ SaaS benchmark, because every inference query burns real compute.

Heroku started killing free dynos on November 28, 2022, and blamed fraud and abuse, an early sign the free tier was breaking.

Reverse trials, usage-based pricing, and BYOK now cap the open-ended compute liability that classic freemium left wide open.

Free access is drifting to the hyperscalers, who can swallow inference costs that would bankrupt a startup, so freemium becomes a monopoly moat.

## REFERENCES

- [1] Midjourney, via The Verge, "Midjourney ends free trials of its AI image generator due to 'extraordinary' abuse." <https://www.theverge.com/2023/3/30/23662940/midjourney-ends-free-trials-ai-image-generator-abuse> <https://www.theverge.com/2023/3/30/23662940/midjourney-ends-free-trials-ai-image-generator-abuse>
- [2] Patrick Campbell, Lenny's Podcast, "On pricing, packaging, and building a bootstrapped B2B empire." <https://www.lennyspodcast.com/patrick-campbell-on-pricing-packaging-and-building-a-bootstrapped-b2b-empire/> <https://www.lennyspodcast.com/videos/10-lessons-on-bootstrapping-a-200m-business-patrick-campbell-profitwell/>
- [3] Kevin Hale, Y Combinator, "Pricing Your SaaS Product." <https://www.ycombinator.com/library/6i-pricing-your-saas-product> <https://www.ycombinator.com/library/6h-startup-pricing-101>
- [4] Bessemer Venture Partners, "State of the Cloud 2023." <https://www.bvp.com/atlas/state-of-the-cloud-2023> <https://www.bvp.com/atlas/state-of-the-cloud-2023>
- [5] Gergely Orosz, "ZIRP and the Software Engineering Industry." <https://blog.pragmaticengineer.com/zirp-and-the-software-engineering-industry/> <https://newsletter.pragmaticengineer.com/p/zirp-software-engineers>
- [6] Heroku, "Heroku's Next Chapter." <https://blog.heroku.com/next-chapter> <https://blog.heroku.com/next-chapter>
- [7] PlanetScale, "PlanetScale is sunseting the Hobby tier." <https://planetscale.com/blog/sunseting-the-hobby-tier> <https://planetscale.com/blog/planetscale-forever>

- [8] Sam Altman, on ChatGPT compute costs. <https://twitter.com/sama/status/1599668808285028353> <https://twitter.com/sama/status/1599668808285028353>
- [9] Sequoia Capital, "Generative AI: A Creative New World." <https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/> <https://sequoiacap.com/article/generative-ai-a-creative-new-world/>
- [10] Sarah Tavel, "Sell work, not software." <https://sarahtavel.substack.com/p/sell-work-not-software> <https://www.sarahtavel.com/p/ai-startups-sell-work-not-software>
- [11] Ben Thompson, Stratechery, "AI and the Big Five." <https://stratechery.com/2023/ai-and-the-big-five/> <https://stratechery.com/2023/ai-and-the-big-five/>
- [12] Baremetrics, "Why We Dropped Our Free Plan." <https://baremetrics.com/blog/why-we-dropped-our-free-plan> <https://baremetrics.com/blog/freemium-saas-implode>
- [13] Elena Verna, Lenny's Newsletter, "The ultimate guide to reverse trials." <https://www.lennysnewsletter.com/p/the-ultimate-guide-to-reverse-trials> <https://www.elenaverna.com/p/reverse-trials-examples>
- [14] Kyle Poyar, "Is freemium dead? Long live the reverse trial." <https://kylepoyar.substack.com/p/freemium-is-dead-long-live-the-reverse> <https://kylepoyar.substack.com/p/your-guide-to-reverse-trials>
- [15] OpenView, "The State of Usage-Based Pricing." <https://openviewpartners.com/blog/state-of-usage-based-pricing/> <https://openviewpartners.com/blog/state-of-usage-based-pricing/>
- [16] Tomasz Tunguz, "The Gross Margins of Generative AI Businesses." <https://tomtunguz.com/generative-ai-gross-margins/>
- [17] TechCrunch, "Heroku announces plans to eliminate free plans, blaming 'fraud and abuse'." <https://techcrunch.com/2022/08/25/heroku-announces-plans-to-eliminate-free-plans-blaming-fraud-and-abuse/> <https://techcrunch.com/2022/08/25/heroku-announces-plans-to-eliminate-free-plans-blaming-fraud-and-abuse/>
- [18] Docker, "We apologize: We did a terrible job announcing the end of Docker Free Teams." <https://www.docker.com/blog/we-apologize-we-did-a-terrible-job-announcing-the-end-of-docker-free-teams/> <https://www.docker.com/blog/we-apologize-we-did-a-terrible-job-announcing-the-end-of-docker-free-teams/>
- [19] GitLab, "GitLab Premium pricing update." <https://about.gitlab.com/blog/2023/03/02/gitlab-premium-pricing-update/> <https://about.gitlab.com/blog/gitlab-premium-update/>