Chutes

Brief 012 - Chutes (Subnet 64)

Supporting information / resources:

Brand assets: Google Drive

How Chutes works

Chutes provides decentralized compute as a service for AI applications to be developed. Deploy, run and scale any AI model in seconds. Try directly through our platform, or use our easy-to-use API.

Decentralized infrastructure powered by Bittensor's Subnet 64, where contributors provide GPU/compute in exchange for Chutes alpha.

Fully **serverless**: deploy any AI model (LLMs, pipelines, custom code) to a global decentralized infrastructure in seconds.

What advantages does Chutes have over centralized competitors

- **Cost-efficiency**: Often ~85% cheaper than AWS/GCP for inference workloads due to a decentralized marketplace

- **Instant provisioning**: No VM setup—just click to deploy and start immediately

- **Scalability & resilience**: Distributed compute avoids single-point failures, and expands with network growth.

- **Token incentives**: GPU providers are rewarded transparently in Chutes Alpha, aligning the entire ecosystem for performance and quality.

- **Open-Source**: so enterprise teams can truly trust our isolated environments (TEE).

Chutes don't have one datacenter, they have dozens. Chutes don't have one IT support team, they have nearly 100.
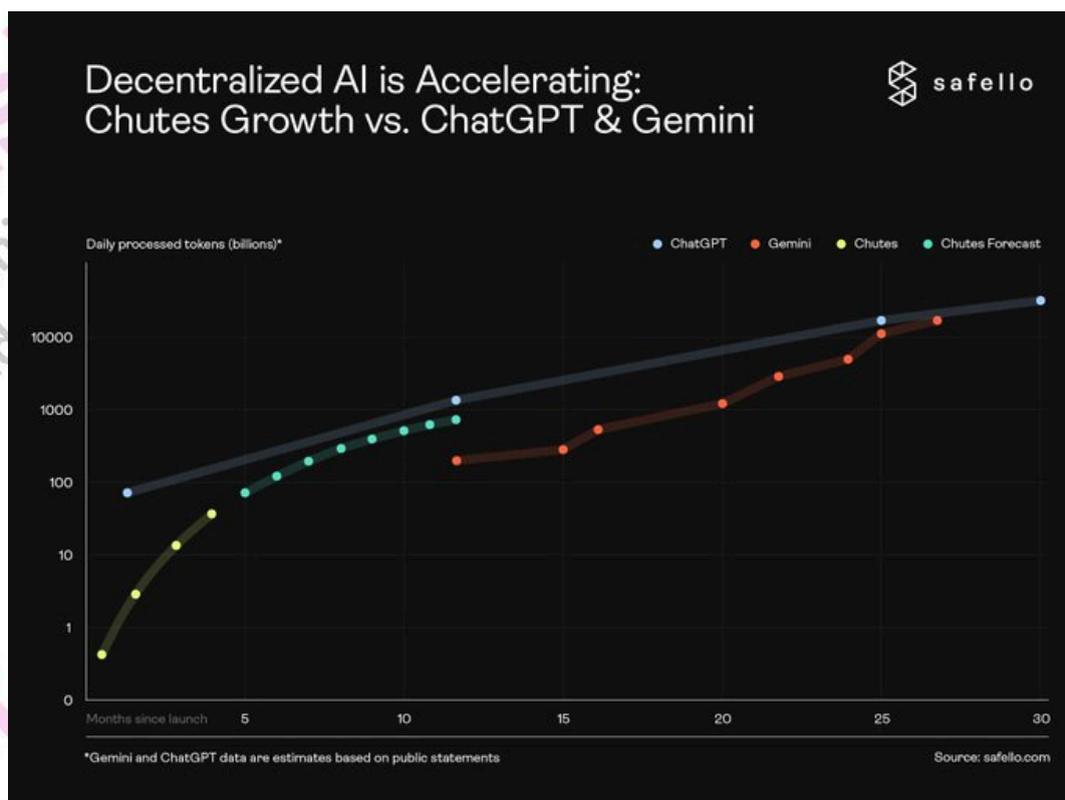
Use cases

Chutes is not just an LLM inference platform, it has many categories of apps/models and custom code options such as:
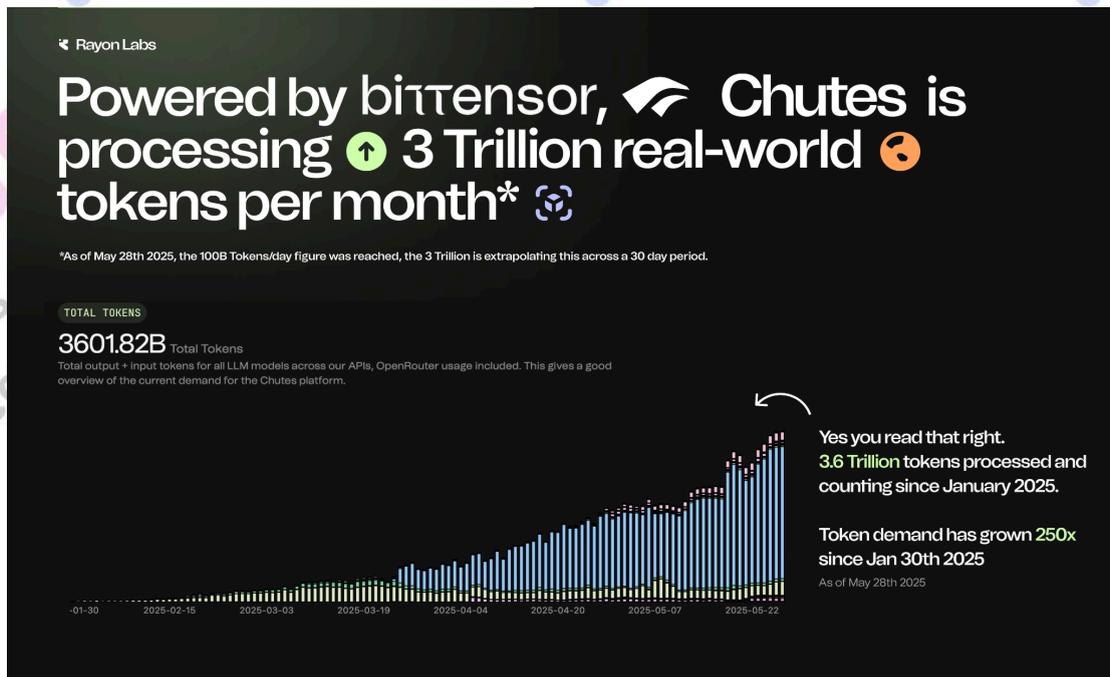
- Text to video
- Image editing
- Rentals
- Model training
- AI Transcription
- Music Generation
- AI Content Moderation
- ChatGPT/Midjourney-style user interfaces
- Custom Models/AI Workflows

Chutes token output

Within just 5 months of launch, Chutes surpassed 100 billion tokens processed in a single day — matching Google's volume from just a year ago.

More recently, Chutes has processed nearly **160 billion tokens daily** — cementing its place as one of the most powerful ai compute layers in crypto.

Subscription tiers for payments/usage

On June 17th Chutes Payments with fiat went Live in Beta, accepting fiat payments for all paid models.

Chute currently run a 'base tier' which includes 200 requests/day globally, with 1/10 of the cost for prompt rerolls. Anything past that is paid on a pay-as-you-go basis with model pricing available on each page of our website. The plan is to add paid subscription tiers too, so users can pay for X usage per month on a recurring basis.

https://x.com/chutes_ai/status/1935037565372953006

Current revenue generation

$61K USD revenue in the last week = ~$3.17M per year when extrapolated. FIAT payments have only launched this month with the income only climbing over time. Of course, as previously stated, **Chutes uses all revenue generated through its platform to perform buybacks**. The recent revenue may be subject to slightly elevated values due to the one-off fees expected for existing users.

There was only a 37.93% drop off of usage after migration to enforcing a minimum payment for models.

Size of the market

AI Inference market is >$106.03 billion USD which is expected to rise to over $255 billion USD in the coming years.

Road map

Chutes are planning to add TEE (trusted execution environments) to enable secure, separate, isolated environments. **It's a huge move for Chutes as it enables onboarding of enterprise customers and as our code is open-source it provides true end-to-end trust that centralised, closed-source providers can't.**

Long Jobs are coming soon: enabling longer and more substantial tasks. An example use-case would be heavy video models that take a long time to generate.

Consumer-focused applications: Chutes Chat is getting a complete makeover with brand new features and platforms. Chutes Studio will arrive soon with image generation tools, video and more.

Further revenue streams: Chutes are adding revenue to consumer apps, further paid tiers for inference, alongside the existing PAYG and Squad revenue sources.

Apps built on Chutes

Squad - a platform for building, deploying, and scaling AI agents.

> Examples: Michael Taolor - Bittensor copilot for subnet trading research and intelligence. (https://x.com/michaeltaolor/status/1923432490083156462)

https://chat.chutes.ai : chatbot powered by Chutes, allowing users to freely interact with any of our large language models.

Coming soon: Chutes Studio, an image/video studio powered by Chutes.

Please contact Chutes for interviews/AMA's/Podcast

Message Chutes on X @chutes_ai

Or go to their discord https://discord.com/invite/HbysyPuZns

Website: https://chutes.ai/

Chart: https://taostats.io/subnets/64/chart