

Forum:[Hack Forums](#)

Thread: [One API, Any AI Model / CHATGPT CLAUDE GEMINI SEEDANCE ++ @ 20-50% OFF API TOKENS](#)

wooha - June 23, 2026, 5:53 am

[BETA OPEN] BeeSwitch — 100+ AI Models / 20+ Providers / 20-50% OFF Open API Token Pricing

Free Pro Plan access until official launch for approved beta users

BeeSwitch is now opening its private beta waiting list.

BeeSwitch is an API-first AI access platform for developers, agencies, automation builders, SaaS teams, and high-volume AI users who do not want to manage separate provider accounts, API keys, invoices, balances, and pricing structures.

Access leading AI models through one BeeSwitch account — with flexible routing, centralized billing, lower-cost token routes, and a platform designed for real API workloads.

This is not a chat UI sold as an API.

BeeSwitch is a multi-model API gateway built around direct integrations, unified access, intelligent routing, usage control, and lower effective token costs.

Who BeeSwitch Is For

BeeSwitch is a strong fit for:

- Developers building AI products
- Latest AI Video Generation model with Gemini, Seedance2 etc
- SaaS founders who need flexible multi-model access
- Automation builders using agents and workflows
- Agencies managing multiple client environments
- Content teams using LLM, image, video, or research models
- AI coding users running Claude Code, Codex, Gemini CLI, or OpenCode
- Prompt engineers and AI power users
- Teams that need lower-cost Token access
- Businesses tired of maintaining multiple API keys, balances, and provider accounts
- Non-Critical uptime access projects
- Companies that require custom quotas, invoices, compliance controls, or higher uptime needs

Why BeeSwitch?

- ? One account for 20+ AI providers
- ? 100+ models through a unified API platform
- ? API-first access for apps, agents, workflows, automations, SaaS products, and internal tools
- ? Competitive pricing on both input and completion/output tokens

- ? No need to maintain separate balances and billing relationships across providers
- ? Switch supported models with minimal integration changes
- ? Multi-route access designed to reduce provider lock-in
- ? Centralized usage visibility, quotas, and billing options
- ? Direct beta onboarding and early-user support

BeeSwitch never uses your prompts, files, or outputs to train any models.

Popular Model Categories

- GPT models
- Claude models, including Opus tiers
- Gemini models
- DeepSeek models
- Seedance-2 and other media / generation models
- Additional leading providers and model routes as availability expands

Built for API Users — Not Just Prompting

BeeSwitch is designed for teams that need more than a single model endpoint.

Use BeeSwitch for:

- Production SaaS features
- AI coding workflows
- Claude Code, Gemini CLI, Codex, OpenCode, and VS Code integrations
- n8n and workflow automations
- AI agents and RAG knowledge bases
- Customer support and internal operations tools
- AI content, research, image, and video generation pipelines
- Multi-client agency environments

For most existing OpenAI-compatible setups, migration is straightforward:

Update your API base URL + API key + model ID, then go live.

No need to rebuild your entire application just to test another supported provider or model route.

Lower API Pricing — Built for Real Usage

BeeSwitch is designed to reduce AI API costs without forcing you into a single-provider ecosystem.

Depending on the model, provider route, and usage level, users can receive:

25%–50% lower pricing versus standard public API list pricing on selected models.

This applies to both:

- Input tokens — prompts, system instructions, uploaded context, and request payloads
- Completion / output tokens — generated responses, reasoning output, code, content, and

structured data

Examples of models where BeeSwitch aims to offer meaningful savings:

GPT-5.5

Input Token Pricing: Up to 25%–50% lower

Completion Token Pricing: Up to 25%–50% lower

Claude Opus

Input Token Pricing: Up to 25%–50% lower

Completion Token Pricing: Up to 25%–50% lower

Gemini

Input Token Pricing: Up to 25%–50% lower

Completion Token Pricing: Up to 25%–50% lower

Seedance-2

Input Token Pricing: Up to 25%–50% lower

Completion Token Pricing: Up to 25%–50% lower

Important: Exact rates vary by model, provider route, token type, usage volume, and market conditions. Current pricing is shown inside BeeSwitch before usage.

No fake “unlimited” promises. No unclear pricing.

The goal is simple: more models, lower effective token cost, and less provider-management overhead.

Why Pricing Can Be Better

BeeSwitch combines multi-channel model supply, centralized purchasing power, intelligent routing, and usage governance.

This helps reduce:

- Fragmented provider billing
- Inefficient model selection
- Unnecessary failed retries
- Invalid Token consumption
- Vendor lock-in
- Wasted spend across teams, projects, and workloads

For larger accounts, BeeSwitch can also support usage auditing by team, project, and model — making AI cost allocation easier to manage.

Routing, Stability & Scale

BeeSwitch is designed for workloads where availability, cost, and latency all matter.

Platform capabilities include:

- ? Unified API access across supported model routes
- ? Intelligent routing based on availability, cost, and latency
- ? Health checks and load balancing
- ? Automatic failover and route degradation where supported
- ? Flexible quotas and model access controls
- ? Token-level usage visibility
- ? Centralized monthly billing options
- ? Enterprise options for dedicated routing, private deployment, and 24/7 operations support

For qualified Enterprise arrangements, BeeSwitch can support dedicated infrastructure, disaster-recovery planning, and availability terms up to 99.95% under a written agreement.

Privacy, Security & Enterprise Controls

BeeSwitch never trains on your prompts, files, or outputs.

For Enterprise requirements, available options may include:

- Tenant isolation
- End-to-end encryption
- Role-based access controls
- Fine-grained permissions
- Usage auditing and operational logs
- Private deployment
- Hybrid-cloud delivery
- Dedicated lines and custom routing
- Custom quotas, billing, and support arrangements

BeeSwitch Plans

1. Starter — Free

For developers and users who want to test BeeSwitch before scaling.

- Free account access
- Entry-level API access
- 10–15% Discount off public API route pricing
- Explore supported models and providers
- API key and account management
- Basic usage visibility
- Suitable for prototyping, testing, and light workloads
- Shared beta capacity

Best for: Individual developers, hobby projects, proof-of-concepts, and users comparing models.

2. Pro — \$35/month annual plan or \$50/month flexible monthly plan

The main plan for active developers, agencies, automation users, and teams running meaningful workloads.

- Broader access to supported providers and models
- 25–35% Discount off public API route pricing
- Higher account capacity than Starter
- Priority access to beta features and new model releases
- Faster support response for active users
- Built for production workloads, client projects, AI content pipelines, and automations
- Priority shared capacity during beta

Best for: Developers, SaaS founders, agencies, AI operators, automation builders, and heavy individual users.

3. Scale — \$150/month annual plan or \$180/month flexible monthly plan

For users running larger volume, more advanced workflows, or multiple client / product environments.

- Everything in Pro
- 35–45% Discount off public API route pricing
- Higher usage capacity
- More advanced account functionality
- Priority consideration for requested providers, models, and capabilities
- Better fit for high-volume automation, internal products, and agency-scale usage
- Enhanced usage and account-management options

Best for: Growing SaaS products, agencies, AI workflow operators, and teams with meaningful monthly API spend.

4. Enterprise — Custom

For companies that need a tailored commercial, technical, security, or uptime setup.

- Custom volume pricing
- Consolidated billing options
- Invoice billing
- Larger-scale account support
- Custom onboarding and integration assistance
- Team and organization-level requirements
- Commercial arrangements for sustained high-volume API usage
- Dedicated routing and line options
- Private deployment or hybrid-cloud options
- 24/7 operations support options
- Contractual Enterprise SLA options, including up to 99.95% availability for qualified deployments

Best for: Businesses, platforms, larger agencies, AI products, and teams with specialized requirements.

Beta Offer

We are accepting a limited number of approved beta users.

Approved beta users receive:

- ? Free Pro Plan access until official launch, plus continued Pro-level token discounts after launch
- ? Access to the plan normally priced at \$35/month on annual billing
- ? Early access to BeeSwitch's provider and model catalog
- ? Pro-level Token API Discounts
- ? Priority consideration for future features, model routes, providers, and integrations
- ? 1-on-1 direct support while the platform is being improved
- ? A chance to influence the roadmap before public launch

There is no charge to join the waiting list.

Beta access is manually reviewed to manage capacity, reduce abuse, and prioritize serious users who can provide practical feedback.

Beta Signup Process

This is currently a manual signup process.

To join the BeeSwitch beta waiting list, reply below or send a direct message with:

1. Your email address
2. Your main use case
3. Whether you need API access for personal, agency, SaaS, automation, or Enterprise use
4. Your approximate monthly AI API spend or expected usage level
5. The providers or models you use most often
6. Any coding tools, frameworks, or workflows you need to connect
7. Any hard uptime, security, billing, or compliance requirements

Do not send API keys or sensitive production credentials.

We will review applications and invite approved users in batches.

BeeSwitch — Switch smarter. Build faster. Spend less.
[contract=tid]Start Contract[/contract]

wooha - July 11, 2026, 11:32 pm

20-50% OFF OPENAI, ANTROPIC, GEMINI, SD API Token pricings!
Run your projects at cost efficiencies unmatched with 1 platform!

wooha - July 20, 2026, 10:02 pm

Newest GPT 5.6 / Fable 5 models have been UP and running SMOOTHLY
20-50% OFF OPENAI, ANTROPIC, GEMINI, SD API Token pricings!
Run your projects at cost efficiencies unmatched with 1 platform!

kannss - July 21, 2026, 3:55 am

i want to use video creation AI only. seedance is preference. what you will charge for this ?

MR.Worldwide - July 22, 2026, 6:06 am

same here, only seedance 2.0 needed. How to access your platform ?

wooha - July 23, 2026, 1:08 am

(Jul 20, 2026 10:02 PM)wooha Wrote: [\[CITE\]](#) Newest GPT 5.6 / Fable 5 models have been UP and running SMOOTHLY

20-50% OFF OPENAI, ANTROPIC, GEMINI, SD API Token pricings!

Run your projects at cost efficiencies unmatched with 1 platform!

(Jul 21, 2026 03:55 AM)kannss Wrote: [\[CITE\]](#) i want to use video creation AI only. seedance is preference. what you will charge for this ?

To get access (PRO PLAN level: \$50/mo free during this beta launch), send me direct message with the below, and we'll get back to you with access:

1. Your email address, telegram/whatsapp
2. Your main use case
3. Whether you need API access for personal, agency, SaaS, automation, or Enterprise use
4. Your approximate monthly AI API spend or expected usage level
5. The providers or models you use most often
6. Any coding tools, frameworks, or workflows you need to connect
7. Any hard uptime, security, billing, or compliance requirements

wooha - July 25, 2026, 4:12 am

20-50% OFF INPUT, OUTPUT and CACHE Tokens for massive savings on your AI projects

Sephiroth - July 25, 2026, 5:11 am

I only want something for coding with ZDR (Zero Data Retention) and only one model is enough for me (GPT 5.6 Sol)

But those prices are kinda expensive? Not actually cheaper?

My current ChatGPT 5.6 subscription is around \$10 a month but only problem is it doesn't have ZDR.

Vedic - July 25, 2026, 5:28 am

Anything for video creation?

wooha - July 29, 2026, 9:54 am

Your AI powered API Projects At MASSIVE savings. Newest GPT 5.6 / Fable 5 models have been UP and running SMOOTHLY

20-50% OFF OPENAI, ANTROPIC, GEMINI, Seedance 2, nano banana, API Token pricings!

Run your projects at cost efficiencies unmatched with 1 platform!

Most popular Text Models / Codex available:

Anthropic Fable 5

GPT 5.6

Most popular Imaging & Video Models available:

Seedance 2

GPT Image 2

Nano banana

blufia118 - August 1, 2026, 3:05 pm

ill try it out, send me info

wooha - August 2, 2026, 8:43 pm

(Aug 1, 2026 03:05 PM)blufia118 Wrote: [\[CITE\]](#) ill try it out, send me info

Dm'd you.

wooha - August 5, 2026, 8:43 am

How to Apply

This is a manual beta signup process. Reply below or send a DM with:

Main use case

Personal, agency, SaaS, automation, or Enterprise use

Approximate monthly API spend or expected volume

Main providers/models you use today

Tools or workflows you need to connect

Any uptime, security, billing, or compliance requirements

Do not send API keys or production credentials.

Current models at significant savings:

gpt-5.5

gpt-5.6-luna

gpt-5.6-sol

gpt-5.6-terra

gpt-image-2

gpt-image-2-all

seedance-2.0-1080p

seedance-2.0-480p

seedance-2.0-720p

seedance-2.0-fast-480p

seedance-2.0-fast-720p

veo3.1-fast-1080p

veo3.1-fast-4k

veo3.1-fast-720p

veo3.1-lite-1080p

veo3.1-lite-4k

veo3.1-lite-720p

veo3.1-quality-1080p

veo3.1-quality-4k

veo3.1-quality-720p

vip-seedance-2.0-fast-480p

vip-seedance-2.0-fast-720p

vip-seedance-2.0-mini-480p

vip-seedance-2.0-mini-720p

vip-seedance-2.0-standard-1080p

vip-seedance-2.0-standard-2160p

vip-seedance-2.0-standard-480p

vip-seedance-2.0-standard-720p

vip-seedance-2-1080p

vip-seedance-2-4k

vip-seedance-2-720p

wooha - August 10, 2026, 10:55 pm

Lots of solutions for MEDIA! Run Your AI powered API Projects At MASSIVE savings.

Media: Seedance 2, nano banana @ up to 50% off.

Run your projects at cost efficiencies unmatched with 1 platform!

Not to mention GPT 5.6, Fable 5, Opus

wooha - August 11, 2026, 11:24 pm

Quote:Do you have Claude api key 200k? Yes we can set up your API key's to be of unlimited token

limits for you - as long as your base setup can run it.

wooha - August 15, 2026, 10:50 pm

Api key's generated and manageable for all your models after you sign up with us!
All models going STRONG with additional 20% bonus savings this month on top-ups.
Get connected NOW.

wooha - August 17, 2026, 10:17 pm

Free voucher copies now available!

wooha - August 20, 2026, 7:06 am

Running best are Media : Seedance 2.0 (Seedance 2.5 upcoming for waitlist!!!), Nano Banana!

wooha - August 24, 2026, 3:32 am

Seedance 2.5 launched and opened for certain whitelisted clients, but in the meanwhile, We're running Seedance 2 at an ADDITIONAL 10% discount IN ADDITION to the current discounts available for current plans to celebrate this milestone!

wooha - August 25, 2026, 9:25 pm

Stable connectivity and 99% uptime across Text models!

wooha - August 27, 2026, 7:26 pm

We've revamped Nano Banana 2 models to be even BETTER VALUE. Get Nano Banana 2 MAX fidelity at a PER GENERATION price the equivalent of NB2 LITE!

wooha - August 30, 2026, 3:07 am

PRIVACY IS KEY

Beeswitch doesn't see, record, nor store ANY of your prompt in / return out requests data. 100% security and anonymity from our part as a passthrough gateway. We don't believe that any gateway should store (even for 1 hour, or 1 second) your private prompts or return requests. How long YOU can keep your own return content is up to your personal setup/provider that you're using.

Long answer: Our gateway is structured at the very root to not access, and not store (not even for 1 hour) any prompt/return data (content, context, images or assets). It goes into destination provider directly (Claude, OpenAI, etc). We 100% believe in content/asset anonymity for ALL of our users. How providers manage though, is up to them. What we DO store is only usage data like model, tokens, duration, request IDs etc.

MassMax - August 31, 2026, 5:27 am

I am a SaaS founder. Would you mind to send me a pro trial?

NewGeek - August 31, 2026, 3:32 pm

Would like to try this if the beta invites are still available.

wooha - September 1, 2026, 7:25 am

(Aug 31, 2026 05:27 AM)MassMax Wrote: [\[CITE\]](#) I am a SaaS founder. Would you mind to send me a pro trial?

I've DM'd

(Aug 31, 2026 03:32 PM)NewGeek Wrote: [\[CITE\]](#) Would like to try this if the beta invites are still available.

I've DM'd

Decoy - September 1, 2026, 9:09 am

send you a pm to trial your service

wooha - September 3, 2026, 4:05 am

(Sep 1, 2026 09:09 AM)Decoy Wrote: [\[CITE\]](#) send you a pm to trial your service
Replied. Please check your email

wooha - September 4, 2026, 11:54 pm

As someone asked "so Beeswitch is cheaper than Open router?"
YES, yes we are :)

wooha - September 6, 2026, 9:05 pm

Have to admit, GPT 6 Astra = God mode

We're offering a limited time offer for MASSIVE discounts of up to 50% off Open Router rates on OpenAI models 5.6 on our Pro plan to celebrate for a limited time only!

wooha - September 8, 2026, 7:56 am

Latest details are updated on-site! [warn]

wooha - September 9, 2026, 10:35 pm

One click integration with the top rated API Relay for quick startup and better management!

OtterLover - September 10, 2026, 5:21 am

I want to try it out

wooha - September 14, 2026, 7:22 am

We now have available direct API tokens via AWSBedrock - DM for details!

(Sep 10, 2026 05:21 AM)OtterLover Wrote: [\[CITE\]](#) I want to try it out
DM'd.

wooha - September 15, 2026, 11:32 pm

Frontier model tokens are on flash sale! Hit us up!

wooha - September 21, 2026, 1:47 am

NOW UPDATED AND AVAILABLE Latest Frontier Models

> CLAUDE MAX - Fable 5 / Sonnet 5 / Opus 4.8
> SEEDANCE 2.5

Tokens at 25-50% off direct and openr prices!

devdanger - September 21, 2026, 9:29 am

interested in. Dm me

wooha - September 22, 2026, 8:07 am

(Sep 21, 2026 09:29 AM)devdanger Wrote: [\[CITE\]](#) interested in. Dm me
Conversation DM'd and PM'd

Are we going to AGI? OpenAI just launched new GPT6 sol and luna models today so we're going through another round of model updates! Stay Tuned!

wooha - September 26, 2026, 10:44 pm

All models revamped and optimized with the launch of GPT 6 Luma / Sol!

wooha - September 28, 2026, 12:36 am

AWS Bedrock for Enterprise uses are AVAILABLE with BeeDirect!

Direct token access via AWSB.

No wait times, No pre-AWS applications or AWS use-case, or quota pre-approvals.

Get your tokens running within 1 day on MAX throttles with frontier Claude models!

wooha - September 29, 2026, 7:39 am

AWS Bedrock for Enterprise uses are AVAILABLE with BeeDirect!

Current remaining access is limited in stock! Frontier models going too fast! [warn]

wooha - October 3, 2026, 6:34 am

Signup directly and Mention HForums via contact for Pro plan upgrade and vouch credits!

NewGeek - October 4, 2026, 2:42 pm

Can you share the details!

wooha - October 5, 2026, 8:05 am

(Oct 4, 2026 02:42 PM)NewGeek Wrote: [\[CITE\]](#) Can you share the details!
DM'd!

wooha - October 8, 2026, 2:53 am

20-50% off public pricing on all OPENAI / ANTROPIC API token pricing

gpt-6-astra
Dynamic Pricing —
OpenAI
-

openai
gpt-pro
gpt-6-luna
Dynamic Pricing —
OpenAI
-

openai
gpt-pro
gpt-6-sol
Dynamic Pricing —
OpenAI
-

openai
gpt-pro
gpt-6.1-sol
Dynamic Pricing —
OpenAI
-

openai
gpt-pro
claude-fable-5
Token-based
Anthropic
-
anthropicopenai
claude-max
claude-opus-4-6
Token-based
Anthropic
-
anthropicopenai
claude-max
claude-opus-4-7
Token-based
Anthropic
-
anthropicopenai
claude-max
claude-opus-4-8
Token-based
Anthropic
-
anthropicopenai
claude-max
claude-sonnet-5
Token-based
Anthropic
-
anthropicopenai
claude-max

wooha - October 9, 2026, 8:25 pm

LATEST UPDATE ON ALL MODELS: All the latest frontier models are NOW available and AT DISCOUNTED RATES!

Fable 5.1
Sonnet 5
Opus 5
Opus 5.5
Gpt-6-astra
Gpt-6-luna
Gpt-6-sol
Gpt-6.1-sol

unusualdd - October 9, 2026, 8:35 pm

Shoot me your tg info

wooha - October 11, 2026, 10:41 pm

(Oct 9, 2026 08:35 PM)unusualdd Wrote: [\[CITE\]](#) Shoot me your tg info
DM Replied!

mughalpat - October 11, 2026, 11:01 pm

Check PM, interested

wooha - October 14, 2026, 8:06 am

(Oct 11, 2026 11:01 PM)mughalpat Wrote: [\[CITE\]](#) Check PM, interested
DM replied and emailed!

NEW SIGNUPS

- 1) Registered on our platform: <https://api.beeswitch.com> for a free account
- 2) Email us with your details for the below and we'll review for eligibility of upgrade to the Pro plan

Email Subject: HF Pro Account Whitelisting Vouch Copy

Email reply to these questions:

1. Your a) email address/username used in account registration, b) forum username, c) whatsapp/tg
2. Your main use case
3. Whether you need API access for personal, agency, SaaS, automation, or Enterprise use
4. Your approximate monthly AI API spend or expected usage level
5. The providers or models you use most often
6. Any coding tools, frameworks, or workflows you need to connect
7. Any hard uptime, security, billing, or compliance requirements

Once we receive the above by email, we'll go ahead review the upgrade request + add the vouch credit for you if approved