ChatGPTNextWeb / **NextChat** Public

Code | Issues 691 | Pull requests 127 | Discussions | Actions | Projects | Security and quality 1

main | **39** Branches | **79** Tags | Go to file | Go to file | Code

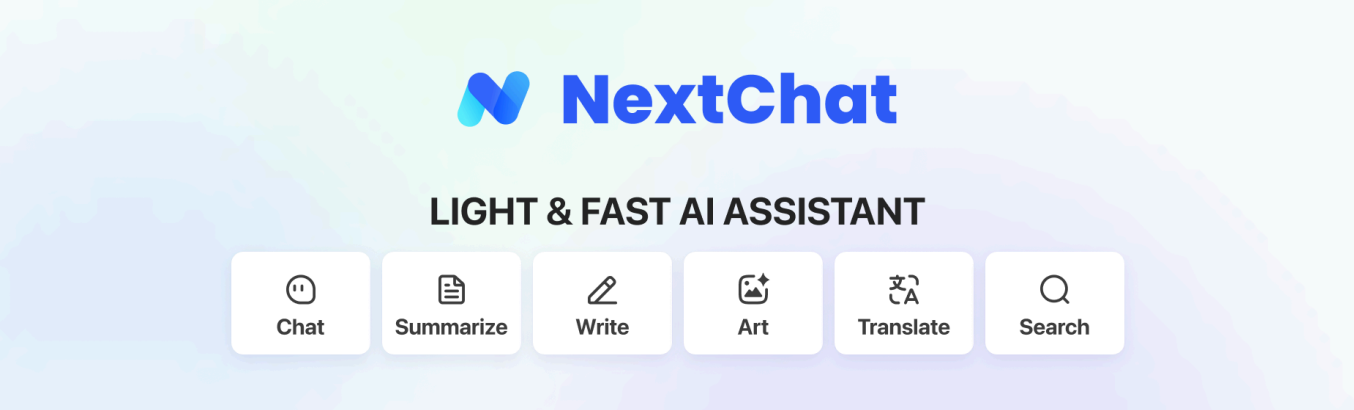
Leizhenpeng Merge pull request #6637 from princeaden1/feat-xai-new-models

c3b8c15 · 8 months ago

.github	update test run target	2 years ago
.husky	feat: add lint-staged	3 years ago
app	feat: new models for xAI (#6559)	8 months ago
docs	docs: Update vercel-ko, cloudflare-pag…	10 months ago
public	Use Vite instead of Create React App	10 months ago
scripts	Add Traditional Chinese prompts conve…	2 years ago
src-tauri	chore: update version	10 months ago
test	test: fix unit test failures	last year
.babelrc	feat: close #2376 add babel polyfill	3 years ago
.dockerignore	fix docker	2 years ago
.env.template	feat: add 302.AI provider	11 months ago
.eslintignore	fix: missing files required for building	last year
.eslintrc.json	chore: add ESLint plugin and rules to r…	2 years ago
.gitignore	feat: ignore mcp_config.json	last year
.gitpod.yml	fix: styles and mobile ux	3 years ago
.lintstagedrc.json	feat: #112 add edit chat title	3 years ago
.prettierrc.js	feat: add lint-staged	3 years ago
CODE_OF_CONDUCT.md	Create CODE_OF_CONDUCT.md	3 years ago
Dockerfile	fix: missing files required for building	last year
LICENSE	Update LICENSE	last year
README.md	docs: update README	10 months ago
README_CN.md	docs: update README	10 months ago
README_JA.md	docs: update README	10 months ago
README_KO.md	docs: Fix typo	10 months ago

docker-compose.yml	chore: update docs for gemini pro	3 years ago
jest.config.ts	test: fix unit test failures	last year
jest.setup.ts	test: fix unit test failures	last year
next.config.mjs	feat: MCP market	last year
package.json	test: fix unit test failures	last year
tsconfig.json	feat: simple MCP example	2 years ago
vercel.json	Update vercel.json	3 years ago
yarn.lock	feat: add 302.AI provider	11 months ago

README | Code of conduct | MIT license



NextChat

English / 简体中文

1 GITHUB TRENDING #1 Repository Of The Day

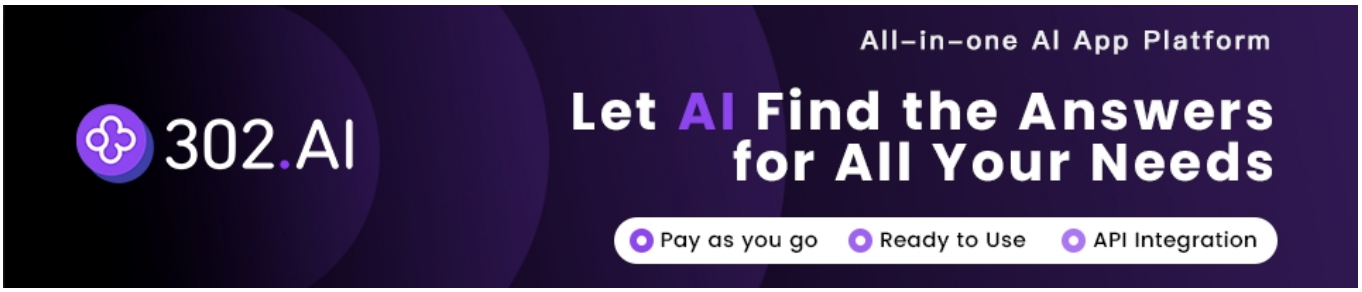
✨ Light and Fast AI Assistant,with Claude, DeepSeek, GPT4 & Gemini Pro support.

NextChat Saas | Web PWA | Windows | MacOS | Linux

NextChatAI / iOS APP / Web App Demo / Desktop App / Enterprise Edition

DEPLOY TO ZEABUR | Deploy | Build with Ona

❤️ Sponsor AI API

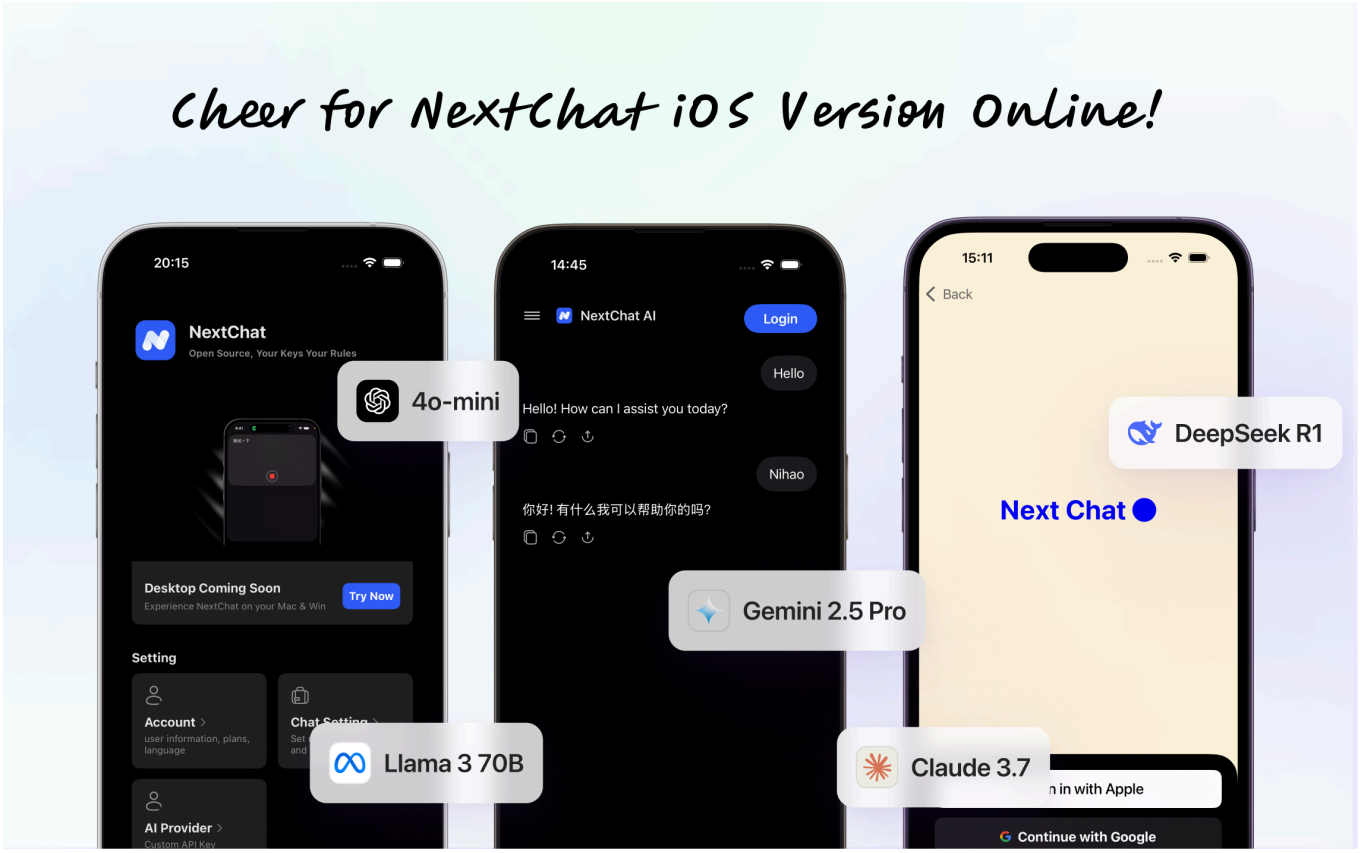


302.AI is a pay-as-you-go AI application platform that offers the most comprehensive AI APIs and online applications available.

🥳 Cheer for NextChat iOS Version Online!

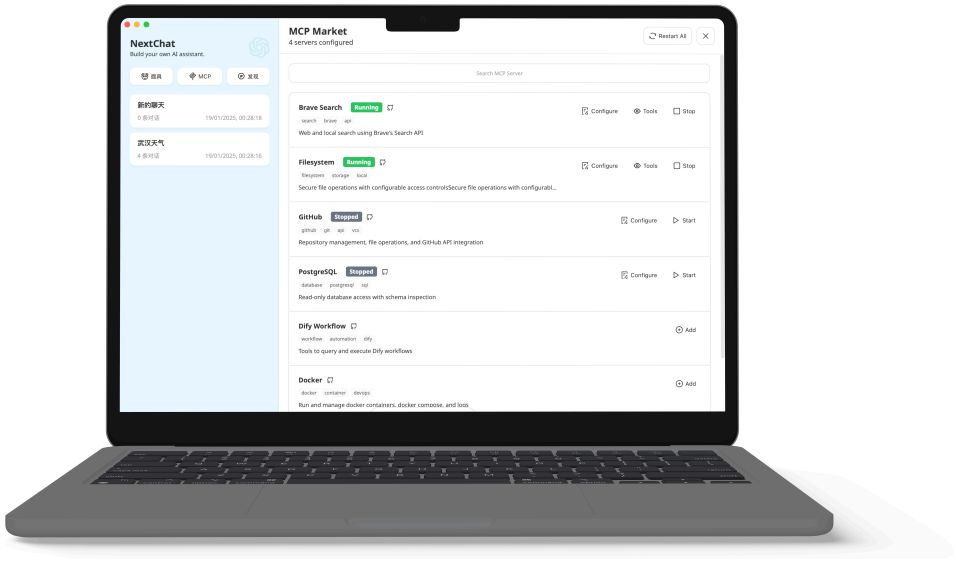
> 👉 Click Here to Install Now

> ❤️ Source Code Coming Soon



🤐 NextChat Support MCP !

> Before build, please set env ENABLE_MCP=true



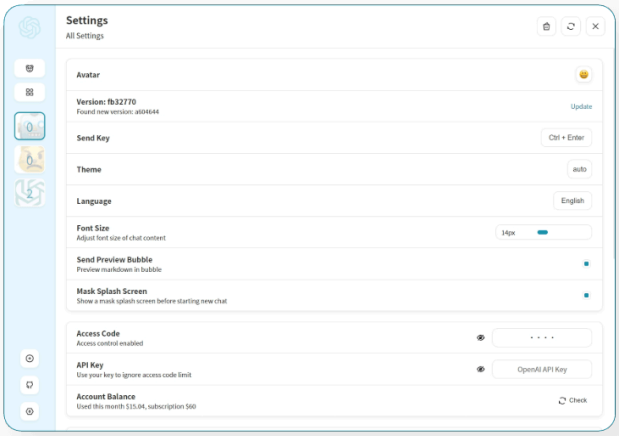
Enterprise Edition

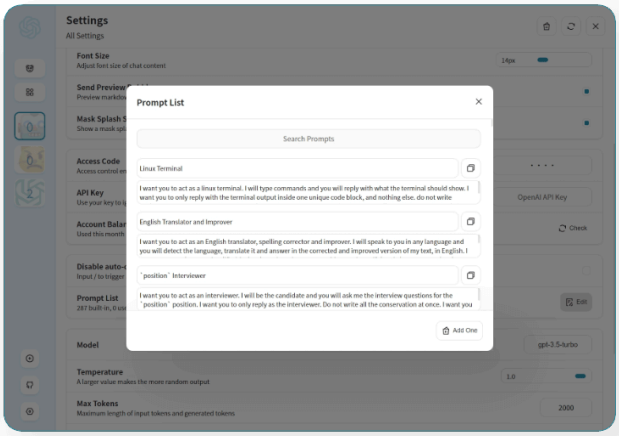
Meeting Your Company's Privatization and Customization Deployment Requirements:

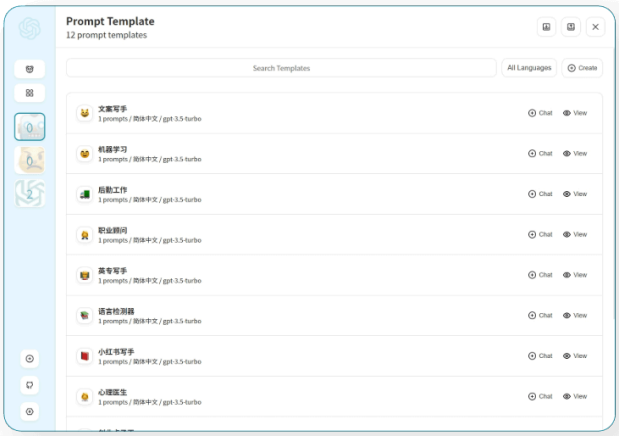
- **Brand Customization**: Tailored VI/UI to seamlessly align with your corporate brand image.
- **Resource Integration**: Unified configuration and management of dozens of AI resources by company administrators, ready for use by team members.
- **Permission Control**: Clearly defined member permissions, resource permissions, and knowledge base permissions, all controlled via a corporate-grade Admin Panel.
- **Knowledge Integration**: Combining your internal knowledge base with AI capabilities, making it more relevant to your company's specific business needs compared to general AI.
- **Security Auditing**: Automatically intercept sensitive inquiries and trace all historical conversation records, ensuring AI adherence to corporate information security standards.
- **Private Deployment**: Enterprise-level private deployment supporting various mainstream private cloud solutions, ensuring data security and privacy protection.
- **Continuous Updates**: Ongoing updates and upgrades in cutting-edge capabilities like multimodal AI, ensuring consistent innovation and advancement.

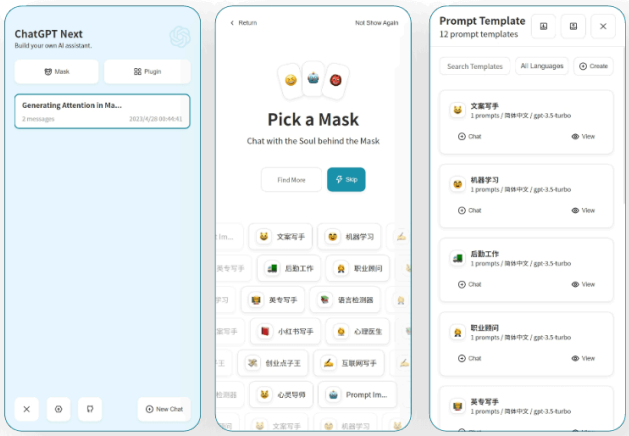
For enterprise inquiries, please contact: **business@nextchat.dev**

Screenshots



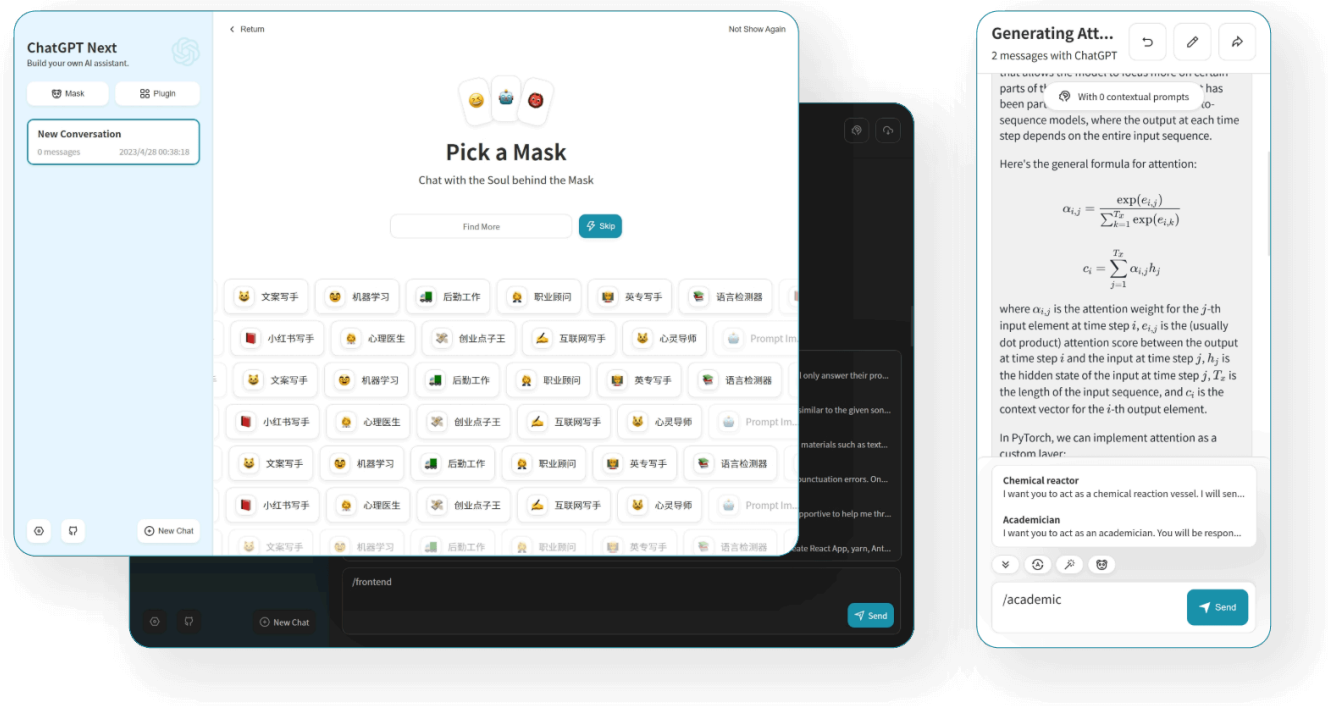






Features

- **Deploy for free with one-click** on Vercel in under 1 minute
- Compact client (~5MB) on Linux/Windows/MacOS, download it now
- Fully compatible with self-deployed LLMs, recommended for use with RWKV-Runner or LocalAI
- Privacy first, all data is stored locally in the browser
- Markdown support: LaTex, mermaid, code highlight, etc.
- Responsive design, dark mode and PWA
- Fast first screen loading speed (~100kb), support streaming response
- New in v2: create, share and debug your chat tools with prompt templates (mask)
- Awesome prompts powered by awesome-chatgpt-prompts-zh and awesome-chatgpt-prompts
- Automatically compresses chat history to support long conversations while also saving your tokens
- I18n: English, 简体中文, 繁体中文, 日本語, Français, Español, Italiano, Türkçe, Deutsch, Tiếng Việt, Русский, Čeština, 한국어, Indonesia



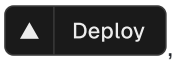
Roadmap

- [x] System Prompt: pin a user defined prompt as system prompt #138
- [x] User Prompt: user can edit and save custom prompts to prompt list
- [x] Prompt Template: create a new chat with pre-defined in-context prompts #993
- [x] Share as image, share to ShareGPT #1741
- [x] Desktop App with tauri
- [x] Self-host Model: Fully compatible with RWKV-Runner, as well as server deployment of LocalAI: llama/gpt4all/rwkv/vicuna/koala/gpt4all-j/cerebras/falcon/dolly etc.
- [x] Artifacts: Easily preview, copy and share generated content/webpages through a separate window #5092
- [x] Plugins: support network search, calculator, any other apis etc. #165 #5353
  - [x] network search, calculator, any other apis etc. #165 #5353
- [x] Supports Realtime Chat #5672
- [ ] local knowledge base

What's New

- 🚀 v2.15.8 Now supports Realtime Chat #5672
- 🚀 v2.15.4 The Application supports using Tauri fetch LLM API, MORE SECURITY! #5379
- 🚀 v2.15.0 Now supports Plugins! Read this: NextChat-Awesome-Plugins
- 🚀 v2.14.0 Now supports Artifacts & SD
- 🚀 v2.10.1 support Google Gemini Pro model.
- 🚀 v2.9.11 you can use azure endpoint now.
- 🚀 v2.8 now we have a client that runs across all platforms!
- 🚀 v2.7 let's share conversations as image, or share to ShareGPT!
- 🚀 v2.0 is released, now you can create prompt templates, turn your ideas into reality! Read this: ChatGPT Prompt Engineering Tips: Zero, One and Few Shot Prompting.

Get Started

1. Get OpenAI API Key;
2. Click 

, remember that `CODE` is your page password;
3. Enjoy :)

FAQ

English > FAQ

Keep Updated

If you have deployed your own project with just one click following the steps above, you may encounter the issue of "Updates Available" constantly showing up. This is because Vercel will create a new project for you by default instead of forking this project, resulting in the inability to detect updates correctly.

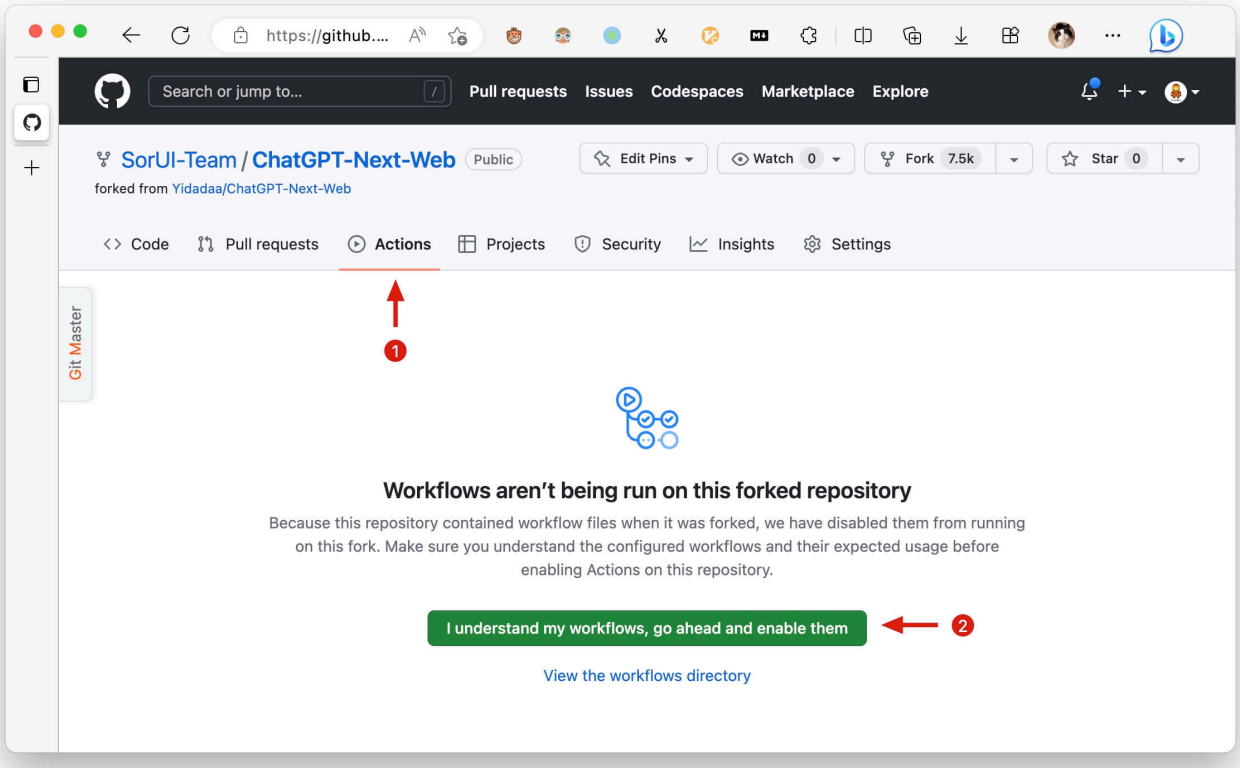
We recommend that you follow the steps below to re-deploy:

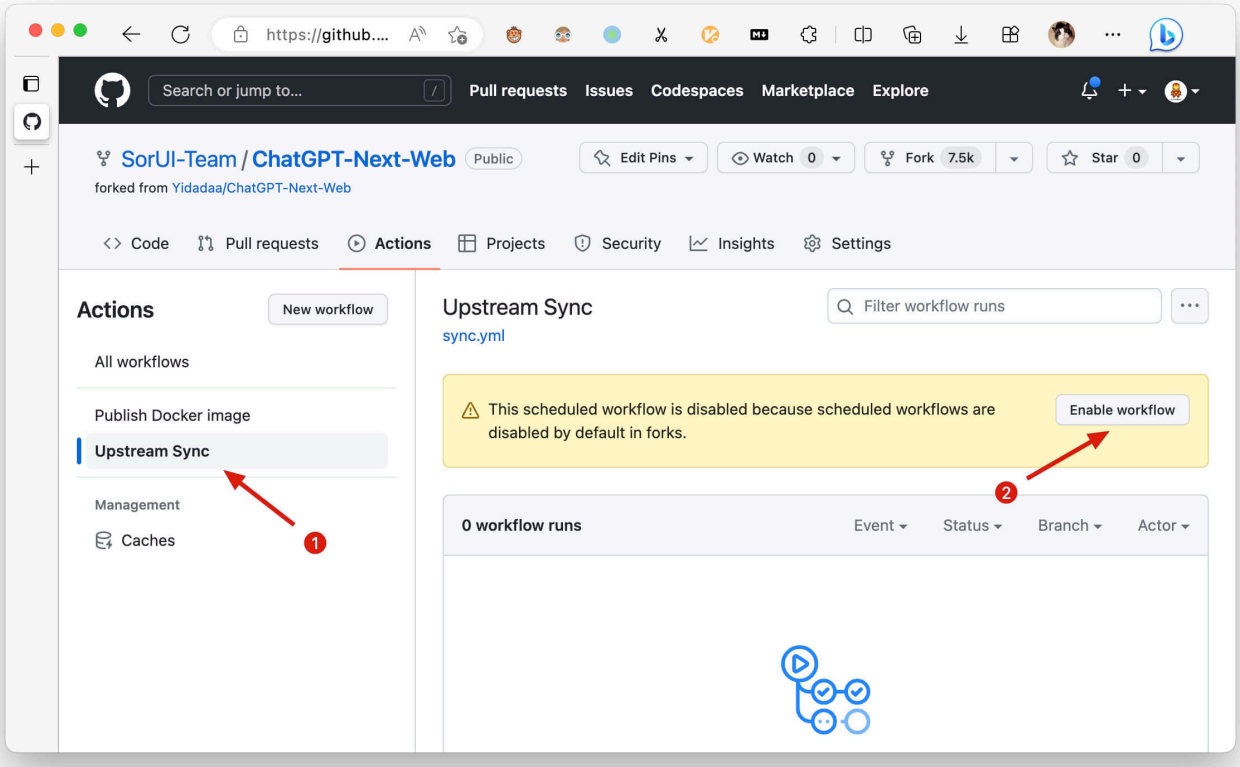
- Delete the original repository;
- Use the fork button in the upper right corner of the page to fork this project;
- Choose and deploy in Vercel again, please see the detailed tutorial.

Enable Automatic Updates

> If you encounter a failure of Upstream Sync execution, please manually update code.

After forking the project, due to the limitations imposed by GitHub, you need to manually enable Workflows and Upstream Sync Action on the Actions page of the forked project. Once enabled, automatic updates will be scheduled every hour:





Manually Updating Code

If you want to update instantly, you can check out the GitHub documentation to learn how to synchronize a forked project with upstream code.

You can star or watch this project or follow author to get release notifications in time.

Access Password

This project provides limited access control. Please add an environment variable named `CODE` on the vercel environment variables page. The value should be passwords separated by comma like this:

```
code1,code2,code3
```

After adding or modifying this environment variable, please redeploy the project for the changes to take effect.

Environment Variables

`CODE` (optional)

Access password, separated by comma.

`OPENAI_API_KEY` (required)

Your openai api key, join multiple api keys with comma.

`BASE_URL` (optional)

> Default: `https://api.openai.com`

> Examples: `http://your-openai-proxy.com`

Override openai api request base url.

`OPENAI_ORG_ID` (optional)

Specify OpenAI organization ID.

`AZURE_URL` (optional)

> Example: https://{azure-resource-url}/openai

Azure deploy url.

`AZURE_API_KEY` (optional)

Azure Api Key.

`AZURE_API_VERSION` (optional)

Azure Api Version, find it at Azure Documentation.

`GOOGLE_API_KEY` (optional)

Google Gemini Pro Api Key.

`GOOGLE_URL` (optional)

Google Gemini Pro Api Url.

`ANTHROPIC_API_KEY` (optional)

anthropic claude Api Key.

`ANTHROPIC_API_VERSION` (optional)

anthropic claude Api version.

`ANTHROPIC_URL` (optional)

anthropic claude Api Url.

`BAIDU_API_KEY` (optional)

Baidu Api Key.

`BAIDU_SECRET_KEY` (optional)

Baidu Secret Key.

`BAIDU_URL` (optional)

Baidu Api Url.

`BYTEDANCE_API_KEY` (optional)

ByteDance Api Key.

`BYTEDANCE_URL` (optional)

ByteDance Api Url.

`ALIBABA_API_KEY` (optional)

Alibaba Cloud Api Key.

`ALIBABA_URL` (optional)

Alibaba Cloud Api Url.

`IFLYTEK_URL` (Optional)

iflytek Api Url.

`IFLYTEK_API_KEY` (Optional)

iflytek Api Key.

`IFLYTEK_API_SECRET` (Optional)

iflytek Api Secret.

`CHATGLM_API_KEY` (optional)

ChatGLM Api Key.

`CHATGLM_URL` (optional)

ChatGLM Api Url.

`DEEPSEEK_API_KEY` (optional)

DeepSeek Api Key.

`DEEPSEEK_URL` (optional)

DeepSeek Api Url.

`HIDE_USER_API_KEY` (optional)

> Default: Empty

If you do not want users to input their own API key, set this value to 1.

`DISABLE_GPT4` (optional)

> Default: Empty

If you do not want users to use GPT-4, set this value to 1.

`ENABLE_BALANCE_QUERY` (optional)

> Default: Empty

If you do want users to query balance, set this value to 1.

`DISABLE_FAST_LINK` (optional)

> Default: Empty

If you want to disable parse settings from url, set this to 1.

`CUSTOM_MODELS` (optional)

> Default: Empty Example: `+llama,+claude-2,-gpt-3.5-turbo,gpt-4-1106-preview=gpt-4-turbo` means add `llama, claude-2` to model list, and remove `gpt-3.5-turbo` from list, and display `gpt-4-1106-preview` as `gpt-4-turbo`.

To control custom models, use `+` to add a custom model, use `-` to hide a model, use `name=displayName` to customize model name, separated by comma.

User `-all` to disable all default models, `+all` to enable all default models.

For Azure: use `modelName@Azure=deploymentName` to customize model name and deployment name.

> Example: `+gpt-3.5-turbo@Azure=gpt35` will show option `gpt35(Azure)` in model list. If you only can use Azure model, `-all,+gpt-3.5-turbo@Azure=gpt35` will `gpt35(Azure)` the only option in model list.

For ByteDance: use `modelName@bytedance=deploymentName` to customize model name and deployment name.

> Example: `+Doubao-lite-4k@bytedance=ep-xxxxx-xxx` will show option `Doubao-lite-4k(ByteDance)` in model list.

`DEFAULT_MODEL` （optional）

Change default model

`VISION_MODELS` (optional)

> Default: Empty Example: `gpt-4-vision,claude-3-opus,my-custom-model` means add vision capabilities to these models in addition to the default pattern matches (which detect models containing keywords like "vision", "claude-3", "gemini-1.5", etc).

Add additional models to have vision capabilities, beyond the default pattern matching. Multiple models should be separated by commas.

`WHITE_WEBDAV_ENDPOINTS` (optional)

You can use this option if you want to increase the number of webdav service addresses you are allowed to access, as required by the format：

- Each address must be a complete endpoint
  > `https://xxxx/yyy`
- Multiple addresses are connected by ', '

`DEFAULT_INPUT_TEMPLATE` (optional)

Customize the default template used to initialize the User Input Preprocessing configuration item in Settings.

`STABILITY_API_KEY` (optional)

Stability API key.

`STABILITY_URL` (optional)

Customize Stability API url.

`ENABLE_MCP` (optional)

Enable MCP（Model Context Protocol）Feature

`SILICONFLOW_API_KEY` (optional)

SiliconFlow API Key.

`SILICONFLOW_URL` (optional)

SiliconFlow API URL.

`AI302_API_KEY` (optional)

302.AI API Key.

`AI302_URL` (optional)

302.AI API URL.

Requirements

NodeJS >= 18, Docker >= 20

Development

Build with Ona

Before starting development, you must create a new `.env.local` file at project root, and place your api key into it:

```
OPENAI_API_KEY=<your api key here>

# if you are not able to access openai service, use this BASE_URL
BASE_URL=https://chatgpt1.nextweb.fun/api/proxy
```

Local Development

```
# 1. install nodejs and yarn first
# 2. config local env vars in `.env.local`
# 3. run
yarn install
yarn dev
```

Deployment

Docker (Recommended)

```
docker pull yidadaa/chatgpt-next-web

docker run -d -p 3000:3000 \
   -e OPENAI_API_KEY=sk-xxxx \
   -e CODE=your-password \
   yidadaa/chatgpt-next-web
```

You can start service behind a proxy:

```
docker run -d -p 3000:3000 \
   -e OPENAI_API_KEY=sk-xxxx \
   -e CODE=your-password \
   -e PROXY_URL=http://localhost:7890 \
   yidadaa/chatgpt-next-web
```

If your proxy needs password, use:

```
-e PROXY_URL="http://127.0.0.1:7890 user pass"
```

If enable MCP, use :

```
docker run -d -p 3000:3000 \
   -e OPENAI_API_KEY=sk-xxxx \
   -e CODE=your-password \
   -e ENABLE_MCP=true \
   yidadaa/chatgpt-next-web
```

Shell

```
bash <(curl -s https://raw.githubusercontent.com/Yidadaa/ChatGPT-Next-Web/main/scripts/setup.sh)
```

Synchronizing Chat Records (UpStash)

| 简体中文 | English | Italiano | 日本語 | 한국어

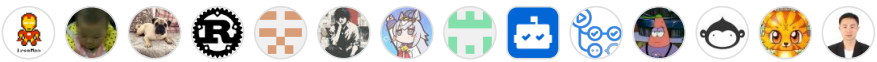
Documentation

Releases 77

v2.16.1 Latest
on Jul 29, 2025

+ 76 releases

Contributors 260



+ 246 contributors

Languages

TypeScript 91.7% **SCSS** 6.9% **Other** 1.4%