

Table of Contents

- [Full / partial translation](#)
 - [Specify source and target languages](#)
 - [Translate with different services](#)
 - [Translate with exceptions](#)
 - [Multi-threads](#)
 - [Custom prompt](#)
 - [Authorization](#)
 - [Custom configuration file](#)
 - [Fonts Subsetting](#)
 - [Translation cache](#)
-

Full / partial translation

- Entire document

```
pdf2zh example.pdf
```

- Part of the document

```
pdf2zh example.pdf -p 1-3,5
```

 [Back to top](#)

Specify source and target languages

See [Google Languages Codes](#), [DeepL Languages Codes](#)

```
pdf2zh example.pdf -li en -lo ja
```

 [Back to top](#)

Translate with different services

We've provided a detailed table on the required [environment variables](#) for each translation service. Make sure to set them before using the respective service.

Translator	Service	Environment Variables	Default Values	Notes
Google (Default)	google	None	N/A	None
Bing	bing	None	N/A	None
DeepL	deepl	DEEPL_AUTH_KEY	[Your Key]	See DeepL
DeepLX	deeplx	DEEPLX_ENDPOINT	https://api.deepl.com/translate	See DeepLX
Ollama	ollama	OLLAMA_HOST, OLLAMA_MODEL	http://127.0.0.1:11434, gemma2	See Ollama
Xinference	xinference	XINFERENCE_HOST, XINFERENCE_MODEL	http://127.0.0.1:9997, gemma-2-it	See Xinference
OpenAI	openai	OPENAI_BASE_URL, OPENAI_API_KEY, OPENAI_MODEL	https://api.openai.com/v1, [Your Key], gpt-4o-mini	See OpenAI
AzureOpenAI	azure-openai	AZURE_OPENAI_BASE_URL, AZURE_OPENAI_API_KEY, AZURE_OPENAI_MODEL	[Your Endpoint], [Your Key], gpt-4o-mini	See Azure OpenAI
Zhipu	zhipu	ZHIPU_API_KEY, ZHIPU_MODEL	[Your Key], glm-4-flash	See Zhipu
ModelScope	ModelScope	MODELSCOPE_API_KEY, MODELSCOPE_MODEL	[Your Key], Qwen/Qwen2.5-Coder-32B-Instruct	See ModelScope
Silicon	silicon	SILICON_API_KEY, SILICON_MODEL	[Your Key], Qwen/Qwen2.5-7B-Instruct	See SiliconCloud
Gemini	gemini	GEMINI_API_KEY, GEMINI_MODEL	[Your Key], gemini-1.5-flash	See Gemini
Azure	azure	AZURE_ENDPOINT, AZURE_API_KEY	https://api.translator.azure.cn, [Your Key]	See Azure
Tencent	tencent	TENCENTCLOUD_SECRET_ID, TENCENTCLOUD_SECRET_KEY	[Your ID], [Your Key]	See Tencent
Dify	dify	DIFY_API_URL, DIFY_API_KEY	[Your DIFY URL], [Your Key]	See Dify , Three variables, lang_out, lang_in, and text, need to be defined in Dify's workflow input.
AnythingLLM	anythingllm	AnythingLLM_URL, AnythingLLM_APIKEY	[Your AnythingLLM URL], [Your Key]	See anything-llm
Argos Translate	argos			See argos-translate
Grok	grok	GORK_API_KEY, GORK_MODEL	[Your GORK_API_KEY], grok-2-1212	See Grok
Groq	groq	GROQ_API_KEY, GROQ_MODEL	[Your GROQ_API_KEY], llama-3-3-70b-versatile	See Grog
DeepSeek	deepseek	DEEPSEEK_API_KEY, DEEPSEEK_MODEL	[Your DEEPSEEK_API_KEY], deepseek-chat	See DeepSeek
OpenAI-Liked	openailiked	OPENAILIKED_BASE_URL, OPENAILIKED_API_KEY, OPENAILIKED_MODEL	url, [Your Key], model name	None
Ali Qwen Translation	qwen-mt	ALI_MODEL, ALI_API_KEY, ALI_DOMAINS	qwen-mt-turbo, [Your Key], scientific paper	Traditional Chinese are not yet supported, it will be translated into Simplified Chinese. More see Qwen MT

For large language models that are compatible with the OpenAI API but not listed in the table above, you can set environment variables using the same method outlined for OpenAI in the table.

Use `-s service` or `-s service:model` to specify service:

```
pdf2zh example.pdf -s openai:gpt-4o-mini
```

Or specify model with environment variables:

```
set OPENAI_MODEL=gpt-4o-mini  
pdf2zh example.pdf -s openai
```

For PowerShell user:

```
$env:OPENAI_MODEL = gpt-4o-mini  
pdf2zh example.pdf -s openai
```

[↑ Back to top](#)

Translate with exceptions

Use regex to specify formula fonts and characters that need to be preserved:

```
pdf2zh example.pdf -f "(CM[RT].*|MS.*|.*Ital)" -c "(\\(|\\||\\)|\\+|=|\\d|[\\u0080-\\ufaff])"
```

Preserve `Latex`, `Mono`, `Code`, `Italic`, `Symbol` and `Math` fonts by default:

```
pdf2zh example.pdf -f "(CM[R]|MS.M|XY|MT|BL|RM|EU|LA|RS|LINE|LCIRCLE|TeX-  
|rsfs|txsy|wasy|stmary|.*Mono|.*Code|.*Ital|.*Sym|.*Math)"
```

[↑ Back to top](#)

Multi-threads

Use `-t` to specify how many threads to use in translation:

```
pdf2zh example.pdf -t 1
```

[↑ Back to top](#)

Custom prompt

Note: System prompt is currently not supported. See [this change](#).

Use `--prompt` to specify which prompt to use in llm:

```
pdf2zh example.pdf --prompt prompt.txt
```

For example:

```
You are a professional, authentic machine translation engine. Only Output the translated
text, do not include any other text.

Translate the following markdown source text to ${lang_out}. Keep the formula notation
{v*} unchanged. Output translation directly without any additional text.

Source Text: ${text}

Translated Text:
```

In custom prompt file, there are three variables can be used.

variables	comment
lang_in	input language
lang_out	output language
text	text need to be translated

 [Back to top](#)

Authorization

Use `--authorized` to specify which user to use Web UI and custom the login page:

```
pdf2zh example.pdf --authorized users.txt auth.html
```

example users.txt

Each line contains two elements, username, and password, separated by a comma.

```
admin,123456
user1,password1
user2,abc123
guest,guest123
test,test123
```

example auth.html

```
<!DOCTYPE html>
<html>
<head>
  <title>Simple HTML</title>
</head>
<body>
  <h1>Hello, World!</h1>
  <p>Welcome to my simple HTML page.</p>
</body>
</html>
```

 [Back to top](#)

Custom configuration file

Use `--config` to specify which file to configure the PDFMathTranslate:

```
pdf2zh example.pdf --config config.json
```

```
pdf2zh -i --config config.json
```

example config.json

```
{
  "USE_MODELSCOPE": "0",
  "PDF2ZH_LANG_FROM": "English",
  "PDF2ZH_LANG_TO": "Simplified Chinese",
  "NOTO_FONT_PATH": "/app/SourceHanSerifCN-Regular.ttf",
  "translators": [
    {
      "name": "deeplx",
      "envs": {
        "DEEPLX_ENDPOINT": "http://localhost:1188/translate/",
        "DEEPLX_ACCESS_TOKEN": null
      }
    },
    {
      "name": "ollama",
      "envs": {
        "OLLAMA_HOST": "http://127.0.0.1:11434",
        "OLLAMA_MODEL": "gemma2"
      }
    }
  ]
}
```

By default, the config file is saved in the `~/.config/PDFMathTranslate/config.json`. The program will start by reading the contents of `config.json`, and after that it will read the contents of the environment variables. When an environment variable is available, the contents of the environment variable are used first and the file is updated.

 [Back to top](#)

Fonts subsetting

By default, PDFMathTranslate uses fonts subsetting to decrease sizes of output files. You can use `--skip-subset-fonts` option to disable fonts subsetting when encountering compatibility issues.

```
pdf2zh example.pdf --skip-subset-fonts
```

 [Back to top](#)

Translation cache

PDFMathTranslate caches translated texts to increase speed and avoid unnecessary API calls for same contents. You can use `--ignore-cache` option to ignore translation cache and force retranslation.

```
pdf2zh example.pdf --ignore-cache
```

 [Back to top](#)

Deployment as a public services

PDFMathTranslate has added the features of **enabling partial services** and **hiding Backend information** in the configuration file. You can enable these by setting `ENABLED_SERVICES` and `HIDDEN_GRADIO_DETAILS` in the configuration file. Among them:

- `ENABLED_SERVICES` allows you to choose to enable only certain options, limiting the number of available services.
- `HIDDEN_GRADIO_DETAILS` will hide the real API_KEY on the web, preventing users from obtaining server-side keys.

A usable configuration is as follows:

```
{
  "USE_MODELSCOPE": "0",
  "translators": [
    {
      "name": "grok",
```

```
    "envs": {
      "GORK_API_KEY": null,
      "GORK_MODEL": "grok-2-1212"
    }
  },
  {
    "name": "openai",
    "envs": {
      "OPENAI_BASE_URL": "https://api.openai.com/v1",
      "OPENAI_API_KEY": "sk-xxxx",
      "OPENAI_MODEL": "gpt-4o-mini"
    }
  }
],
"ENABLED_SERVICES": [
  "OpenAI",
  "Grok"
],
"HIDDEN_GRADIO_DETAILS": true,
"PDF2ZH_LANG_FROM": "English",
"PDF2ZH_LANG_TO": "Simplified Chinese",
"NOTO_FONT_PATH": "/app/SourceHanSerifCN-Regular.ttf"
}
```

 [Back to top](#)