

AI TOOL

A Tasting of NMPA Submission Documents AI Translation

Derek Li

May 2026

AGENDA

Overview

A specialized AI-powered tool for high-fidelity English-to-Chinese NMPA document translation.

01

Application Overview

Features, capabilities & supported formats

02

System Setup

Initialization, authentication & interface

03

Core Features

Glossary, prompts & translation workflows

04

Operations & Support

Troubleshooting & best practices

01

Application Overview

RAG + Fine-tuned Qwen-7B for NMPA-grade translation accuracy.

Application Features

A specialized AI-powered tool for high-fidelity English-to-Chinese NMPA submission document translation.

1

Glossary Support

Enforce terminology consistency across projects with managed glossary sets.

2

Vector Search

Context-aware RAG retrieval for accurate, domain-specific phrasing.

3

Abbreviation Handling

Automated expansion and management of technical acronyms.

4

Prompt Customization

Fine-tune the AI persona and instruction set per project.

5

Supported Formats

DOCX (Word), XML (Structured Data), and Plain Text files.

System Setup

Initialization, authentication & system
verification.

02

System Initialization

1

Launch Application

```
streamlit run  
frontend/app.py
```

Start the Streamlit UI via command line. The system initializes and the browser interface opens automatically.

2

Azure Authentication

```
Complete first-time login
```

Complete the first-time Azure authentication sequence to establish credentials for API and storage access.

3

Verify API Access

```
Check vector storage  
connection
```

Confirm that API keys are valid and the vector storage connection is active before proceeding.

Prerequisites

- ✓ Python Environment
- ✓ Azure Credentials
- ✓ Network Access

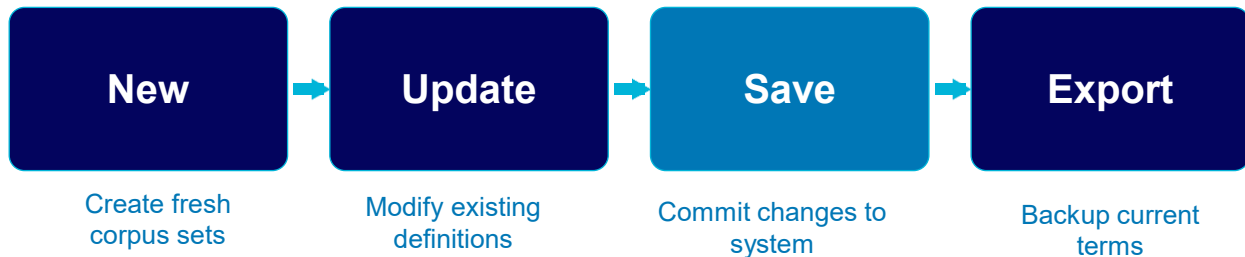
03

Core Features

- Corpus Management
- Prompt Editor
- Translation Workflows

Corpus Management

The Corpus Lifecycle



Required Data Formats

JSON Schema

```
{
  "Adverse Events": "不良事件",
  "Demographics": "人口学",
  "Functional Tests": "功能检查"
}
```

CSV Structure

English	Chinese
Functional Tests	功能检查
Genomics Findings	基因组学研究结果
Adverse Events	不良事件

Corpus Management

System Status

✓ Environment Ready

> Environment Variables

🔄 Refresh Environment

Cache Management

> Cache Status

🗑️ Clear All Caches

Navigation

Settings



Abbreviation List

Available Abbreviation Files

Select an abbreviation file: ⓘ

🔴 abbreviations.json

📁 abbreviations.json: 5 terms

Actions

📄 New Abbreviation List

📄 Export Abbreviation List

🔄 Update Abbreviation List

Vector Library Management

Available Vector Collections

Select a collection: ⓘ

🔴 vector_store_collection_collection

📁 Collection: vector_store_collection_collection

Actions

📄 New Vector Library

📄 Export Vector Library

🔄 Update Vector Library

Translation Dictionary Management

Available Translation Dictionary Files

📁 No dictionary files found. Upload a file to create a new translation dictionary.

Actions

📁 Load Dictionary

Prompt Editor Configuration

Customize the instructional logic sent to the AI model.

Modification Process



Select

Choose the specific prompt template from the available library.

Edit

Modify text instructions directly in the editor interface.

Save

Apply changes to the active translator session.

Configuration Management



Import Config

Load saved configuration files



Export Config

Save current settings externally



Restore Defaults

Reset to factory prompt settings



Safety Protocol — Restore Defaults

Revert to original system instructions if errors occur during prompt editing.

Prompt Editor Configuration

 **System Status**

 Environment Ready

> Environment Variables

 Refresh Environment

 **Cache Management**

> Cache Status

 Clear All Caches

Navigation

Prompt Editor|

 **Prompt Actions**

 Config file exists

Prompt Editor

 Prompt Manager initialized successfully

Edit Prompts

Select a prompt to edit:

system_prompt

Prompt: `system_prompt`

Characters

Description: 系统提示词 - 用于翻译工具和链式翻译

2161

Content - system_prompt

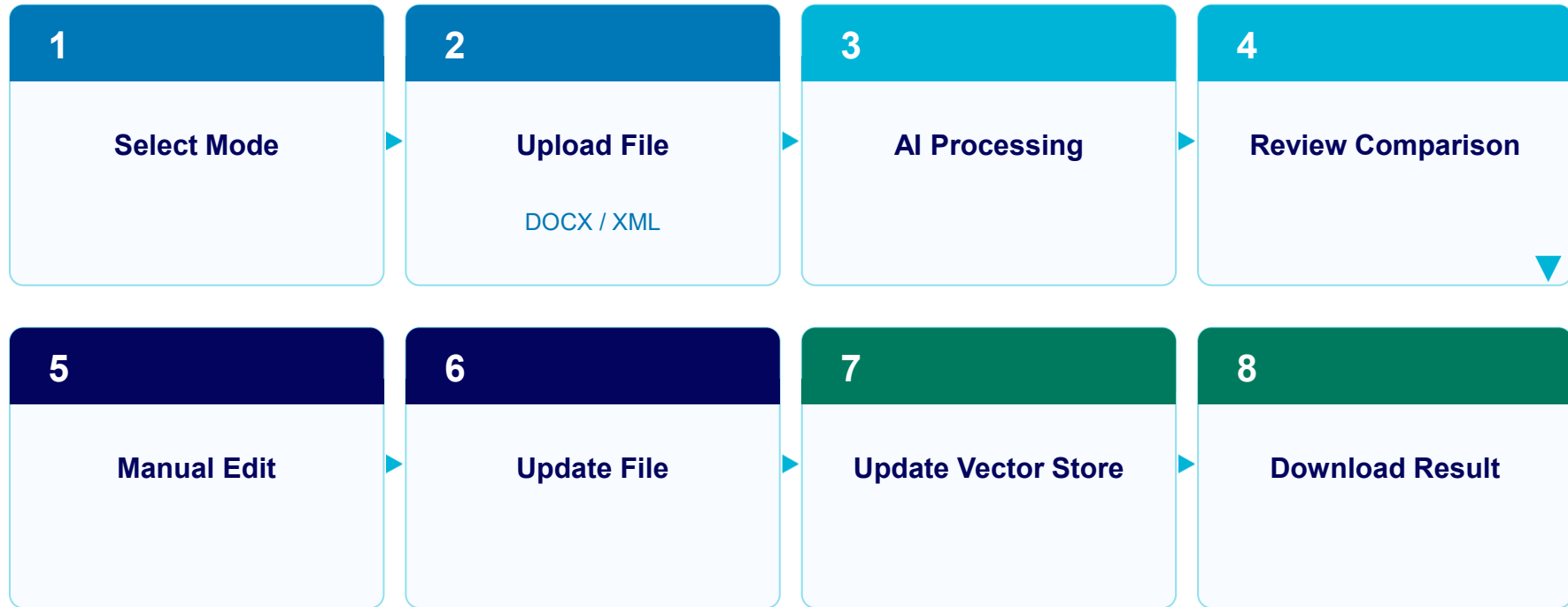
Prompt Content:

You are a professional medical document translation expert. Please translate according to the provided reference documents, maintaining the accuracy and consistency of medical terminology. If there is no relevant content in the reference documents, translate based on your professional knowledge. Reference document content:

{reference_context}{glossary_instruction}

Translation requirements: 1. Maintain the accuracy of medical terminology 2. Preserve the integrity of the meaning 3. Use formal medical

Document Translation Workflow



★ **Critical — Step 7:**

Feeds manual corrections back into the vector store, improving future translation accuracy over time.

Document Translation Workflow

System Status

Environment Ready

Environment Variables

Refresh Environment

Cache Management

Cache Status

Clear All Caches

Navigation

Translation Window

Translation Agent Control

Agent Ready

Reset Agent

Cache Management

Cache Status

Clear All Caches

Navigation

Translation Window

Translation Agent Control

Deploy

File Upload

Upload a file to translate

Drag and drop file here
Limit 200MB per file • DOCX, XML

Browse files

test3.docx 17.0KB

Translation Comparison

Original Text

[1] To demonstrate non-inferior efficacy of gadoquatrane compared to SoC GBCAs1

[2] Using the FAS1 population, the same analytical approach as described in Section 4.2.1.2 will be used; the only difference being the treatment variable consisting of combined pre- and post-gadoquatrane MR image sets vs pre-gadoquatrane image sets.

[3] Abbreviation

[4] Definition

[5] ENR

[6] Enrolled analysis set

Translated Text

[1] 证明gadoquatrane与标准护理 GBCAs__PLACEHOLDER_SUPERSCRIPT_0000__相比具有非劣效性。

[2] 使用FAS1人群，将采用与__PLACEHOLDER_HYPERLINK_0000__部分中描述的相同分析方法；唯一的区别是治疗变量由联合的前后 gadolquatrane MR图像集与前gadolquatrane图像集组成。

[3] 缩写

[4] 定义

[5] ENR

[6] 入组分析集

[7] FAS

Download Translation

Format preserved: DOCX → DOCX

Download DOCX Translation

File: translated_test3.docx (14,997 bytes)

Update Translation File

Update Vector Store

Document Translation Workflow

- The cumulative incidence risk for the event-of-interest, i.e. the proportion of participants with an event of interest over the course of time taking competing risk into account, and, if applicable, the cumulative incidence risk for the competing risk will be estimated for time-to-event variables using **Aalen-Johansen** estimators by study intervention. Estimates will be displayed using plots and complementary tables. The plots of cumulative incidence rates will be truncated at the time point when at least 10% of participants are still under risk [14]. This ensures that any visual signs of treatment differences do not look more extreme than they really are. The Aalen-Johansen estimator is a non-parametric estimator based on the cause-specific Nelson-Aalen estimator and the Kaplan-Meier (KM) estimator $\hat{S}(t)$ and is defined as:

$$\widehat{AJ}_E(t) = \sum_{t_j \leq t} \hat{S}(t_{j-1})(\hat{\Lambda}_E(t_j) - \hat{\Lambda}_E(t_{j-1})).$$

- 对于感兴趣事件的累积发生风险，即在考虑竞争风险的情况下，随时间推移发生感兴趣事件的受试者比例，以及（如适用）竞争风险的累积发生风险，将按研究干预措施采用 **Aalen-Johansen** 估计量对生存时间变量进行估计。估计结果将通过图形和补充表格进行展示。累积发生率曲线将在至少仍有 10% 受试者处于风险中的时间点进行截断 [14]。此举可确保任何治疗差异的视觉表现不会看起来比实际情况更为极端。

Aalen-Johansen 估计量是一种基于原因特异性 Nelson-Aalen 估计量和 Kaplan-Meier (KM) 估计量的非参数估计量 $\hat{S}(t)$ 并被定义为：

$$\widehat{AJ}_E(t) = \sum_{t_j \leq t} \hat{S}(t_{j-1})(\hat{\Lambda}_E(t_j) - \hat{\Lambda}_E(t_{j-1})).$$

Document Translation Workflow

	Day 1	Wk 4	Wk 8	Wk 12	Wk 16	Wk 20	Wk 24	Wk 28	Wk 32	Wk 36	Wk 40	Wk 44	Wk 48 PE	Wk 52	Wk 56	Wk 60
2q8	X	X	X	X	X	O	X	O	X	O	X	O	X	O	X	O
HDq16	X	X	X	O	O	O	X	O	O	O	X	O	O	O	X	O

X – Active Injection, O – Sham Injection, / – No Injection

Primary EP

End of Study



Dose Regimen Modifications (DRM)*

- At Weeks 16 or 20, patients meeting DRM criteria will have treatments scheduled **Q8**
- At Week 24, patients meeting DRM criteria will have treatments scheduled **Q12**
- At subsequent dosing visits after Week 24, participants meeting DRM criteria will have dosing intervals shortened by 4 weeks
- Minimum interval after initiation phase is Q8 for all participants

DRM Criteria for Shortening Dosing Interval:

>5-letter loss in BCVA from Week 12 BCVA due to persistent or worsening DME

AND

>50-micron increase in CST from Week 12

*DRM are applicable only to participants randomized to the HDq16 arm. Participants randomized to 2q8 arm will continue to receive 2 mg aflibercept at 8 week intervals during the entire study treatment period even if observed to meet DRM criteria.

主要终点

研究结束

	第1天	第4周	第8周	第12周	第16周	第20周	第24周	第28周	第32周	第36周	第40周	第44周	第48周 PE	第52周	第56周	第60周
2q8	X	X	X	X	X	O	X	O	X	O	X	O	X	O	X	O
HDq16	X	X	X	O	O	O	X	O	O	O	X	O	O	O	X	O

剂量方案调整 (DRM)*

- 在第16或20周, 符合DRM标准的患者将安排**Q8**治疗
- 在第24周, 符合DRM标准的患者将安排**Q12**治疗
- 在第24周后的后续给药访视中, 符合DRM标准的参与者的给药间隔将缩短4周
- 启动阶段后的最短间隔为所有参与者的Q8

缩短给药间隔的DRM标准:

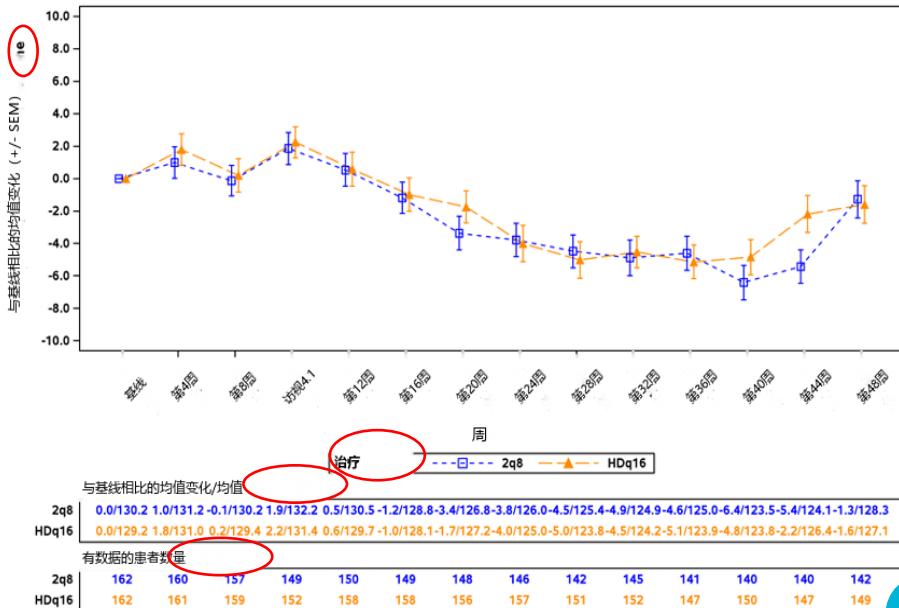
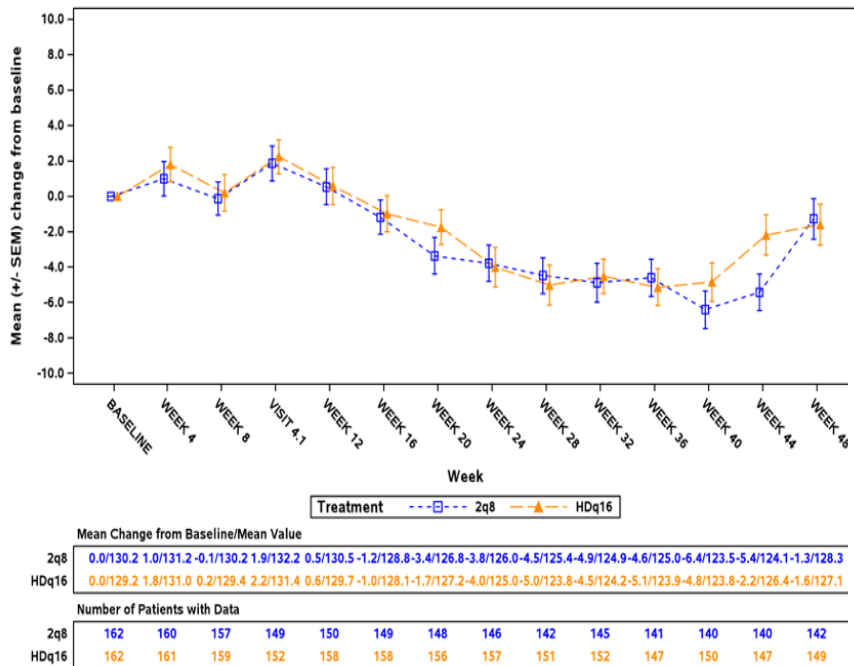
由于持续性或加重性DME导致第12周BCVA的BCVA字母数损失>5个字母

且

从第12周起CST增加>50微米

*DRM仅适用于被随机分配到HDq16组的参与者。被随机分配到2q8组的参与者在整个研究治疗期间以8周间隔继续接受2 mg阿司匹林治疗, 即使观察到符合DRM标准也不例外。

Document Translation Workflow



Operations & Support

Troubleshooting diagnostics & optimization best practices.

A large, light blue circle with a darker blue border, partially visible on the right side of the slide.

04

Troubleshooting & Diagnostics

Translation Failures



Inspect File Formats (Corrupt DOCX/XML)



Check content for unsupported characters



Verify glossary has no conflicting entries

Connection Errors



Check proxy settings



Verify API key validity



Confirm network status and firewall rules

System Health



Cache Issues: Use sidebar to clear cache



API Auth: Perform token refresh



Restart Streamlit service if UI is unresponsive

Optimization & Best Practices

01

PRO TIP

File Hygiene

Ensure clean formatting in DOCX/XML inputs. Remove hidden control characters to prevent parsing errors before upload.

02

PRO TIP

Corpus Strategy

Maintain strict consistency. Avoid duplicate entries with conflicting definitions to prevent model hallucination.

03

PRO TIP

Performance Tuning

Batch similar documents to maintain vector consistency. Clear cache between distinct projects to prevent context bleed.