

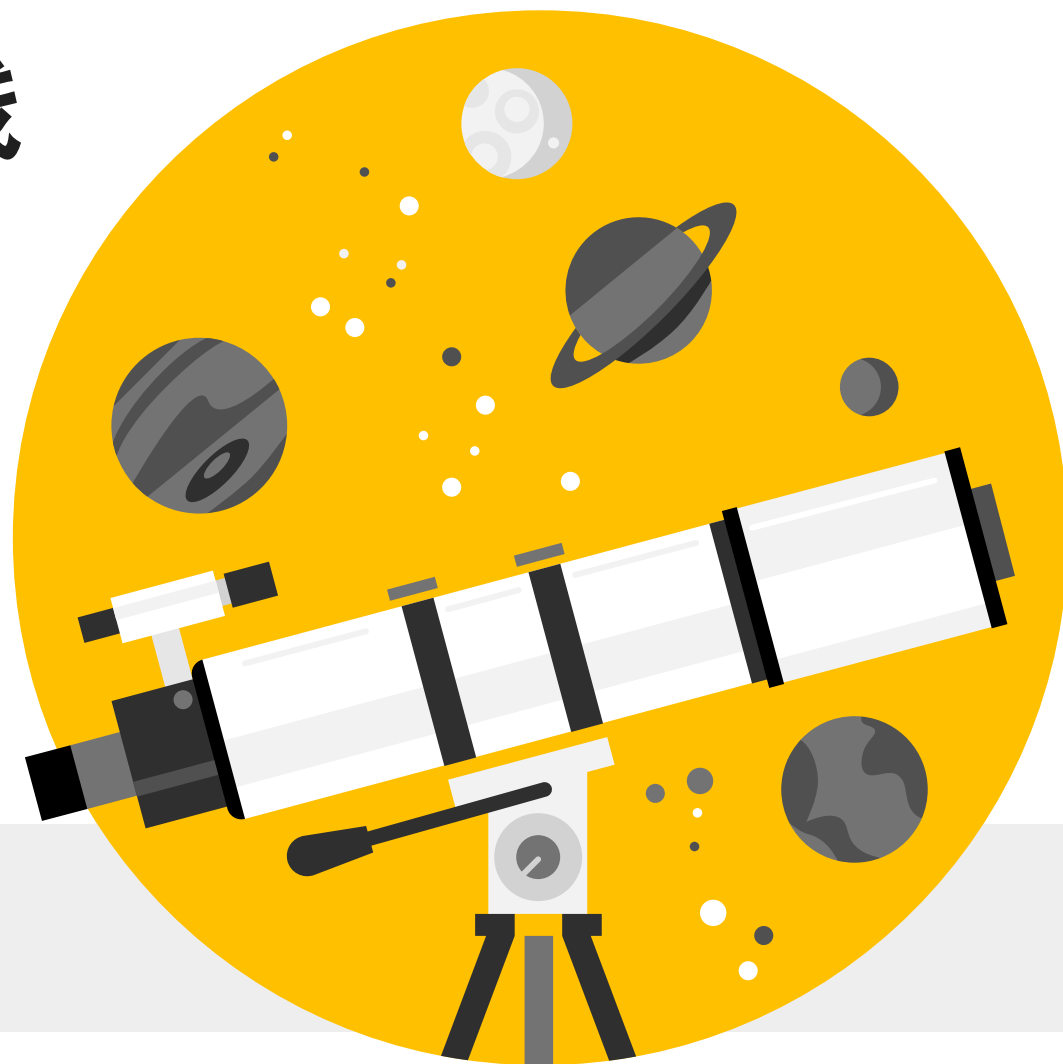
解码译者思维：AI 赋能翻译认知过程 分析与可视化的研究与实践

黄婕

对外经济贸易大学

2025-09-13

北外人工智能翻译研究创新发展高端论坛



汇报内容



1. 翻译认知过程研究

2. 开发自动化的 翻译认知研究平台



3. AI与翻译研究 的交叉创新

4. 译脑智图平台架构 与功能模块



5. 平台初步实现 与技术路线

6. 小结





1. 翻译认知过程研究

翻译认知过程研究的方法

01

眼动追踪

记录译者阅读源文本和产出译文过程中的眼球活动，包括注视点、注视时长和注视顺序等。

02

键盘记录

捕捉译者在翻译过程中的所有键盘操作和暂停，为分析译者工作效率和习惯提供数据支持。

03

脑电图

监测翻译过程中的脑电活动，识别认知负荷的变化，为评估翻译过程提供生理指标。

04

屏幕录制

记录翻译过程中屏幕上的所有活动，包括译者与翻译工具的交互、屏幕变化等。

05

有声思维

要求译者在翻译过程中大声说出自己的思考过程，收集译者认知过程中的语言数据。

翻译认知过程研究的现状与挑战

技术门槛高

眼动研究需要专业设备和技术知识，眼动仪价格昂贵，操作复杂，且数据处理需要专业软件和统计知识。

分析工具缺乏

市场上缺乏专门为翻译认知研究设计的成熟数据分析工具，现有工具通常需要较高的编程技能，且针对性不强。



数据处理复杂

眼动、键盘和脑电等多模态数据的同步与整合极为复杂，研究者需要处理不同格式、不同时间维度的数据。

结果可比性差

不同研究使用不同的数据收集和处理方法，导致研究结果难以进行有效比较，阻碍了领域知识的积累和理论的发展。

2. 开发自动化的 翻译认知研究平台



现有平台的局限性

CRITT TPR-DB介绍

Michael Carl教授的CRITT TPR-DB平台，包含Translog-11、CRITT TPR-DB和YAWAT，是该领域的重要工具。

局限

存在硬件兼容性有限、指标输出有限、学习成本高、缺乏持续维护、自动化程度低和缺乏AI辅助功能等显著局限。



开发新平台的必要性

拓展分析维度

引入更丰富的分析指标和可视化方法，支持更深入的认知过程研究。

提高数据处理效率

利用AI技术自动化数据清洗、同步和分析过程，大幅提升研究效率。

促进标准化

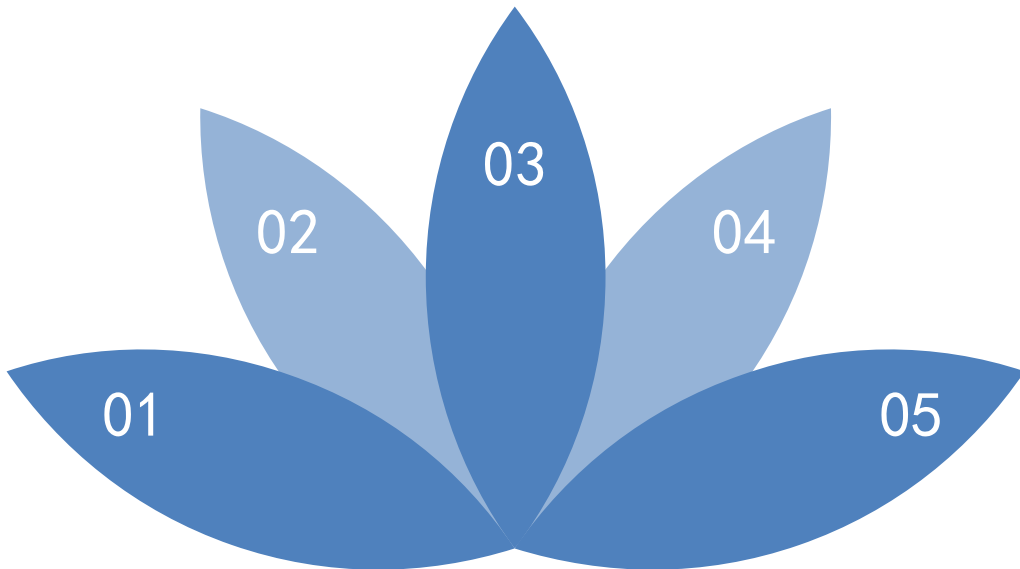
建立统一的数据收集、处理和分析标准，提高研究结果的可比性。

降低技术门槛

通过用户友好的界面和流程化设计，降低翻译研究者进入这一领域的技术障碍。

支持教学应用

将研究成果直观转化为教学工具，促进翻译教育创新。





3. AI与翻译研究 的交叉创新

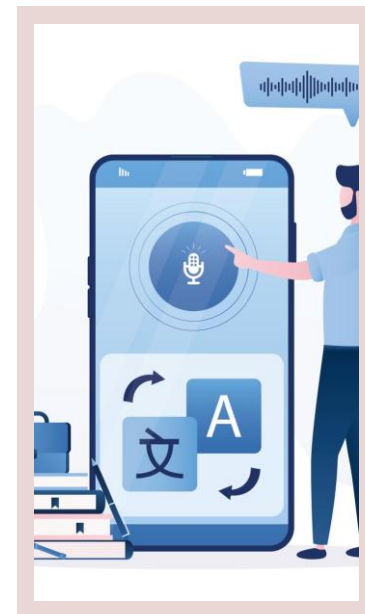
AI辅助实验设计与文本筛选

01 AI评估文本难度

利用大型语言模型（LLMs）进行文本特征分析，自动评估文本词汇密度、句法复杂度，客观量化文本难度，为实验设计提供精准支持。

02 智能识别翻译难点

AI模型能够识别潜在的翻译难点与认知挑战点，如术语定义、复杂句式等，为译者提供预警，助力翻译过程研究与优化。

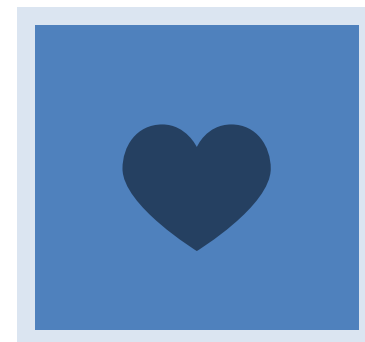
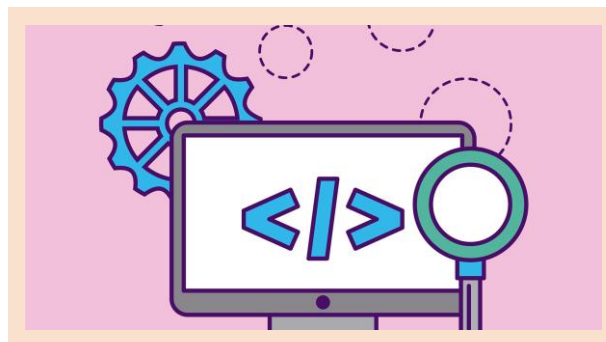


03 精准筛选实验材料

基于指定参数，如词汇难度、句法复杂度等，智能推荐匹配实验目标的文本，确保实验材料的适宜性和代表性。

04 提示工程策略

设计提示工程策略，引导模型识别引发译者认知负荷波动的文本片段，为实验材料的精准筛选提供科学依据，助力翻译过程研究的高效推进。



多模态数据处理的AI赋能

多源数据同步

基于时间戳的精确对齐技术，实现眼动追踪、键盘操作等多源数据的自动同步，确保翻译过程研究的数据一致性。

异常数据过滤

AI算法智能识别并过滤异常数据，如设备故障、操作失误等，提升数据清洗的效率和准确性，为翻译认知研究提供可靠数据支持。

兴趣区划分

自动划分并批量定义兴趣区（AOI），结合翻译过程研究需求，精准捕捉译者视线轨迹和认知负荷变化，深化对翻译过程的理解。



认知过程的AI解析与可视化

机器学习识模式

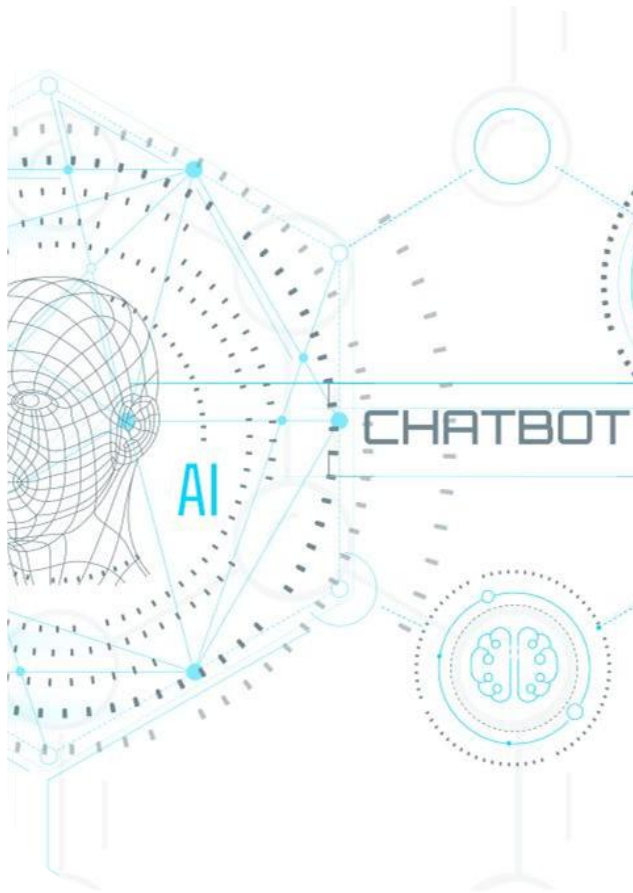
通过先进的机器学习算法，精准识别译者的决策模式，揭示翻译过程中的认知规律，为翻译教学和研究提供新视角。

构建认知负荷模型

在深入理解译者认知过程的基础上，构建翻译认知负荷预测模型，预测不同情境下译者的认知负荷变化，优化翻译策略。

可视化认知过程

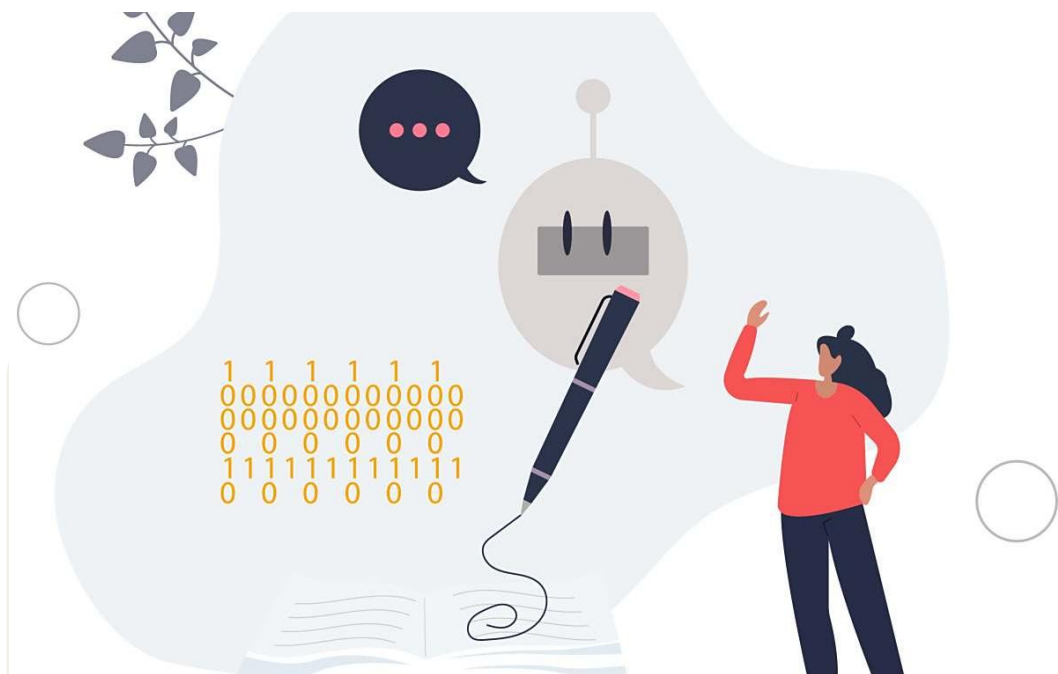
生成交互式的认知过程可视化图表，直观展示译者在面对不同翻译单位时的注意力分配、停留时间与认知负荷变化。



4. 译脑智图平台架构 与功能模块

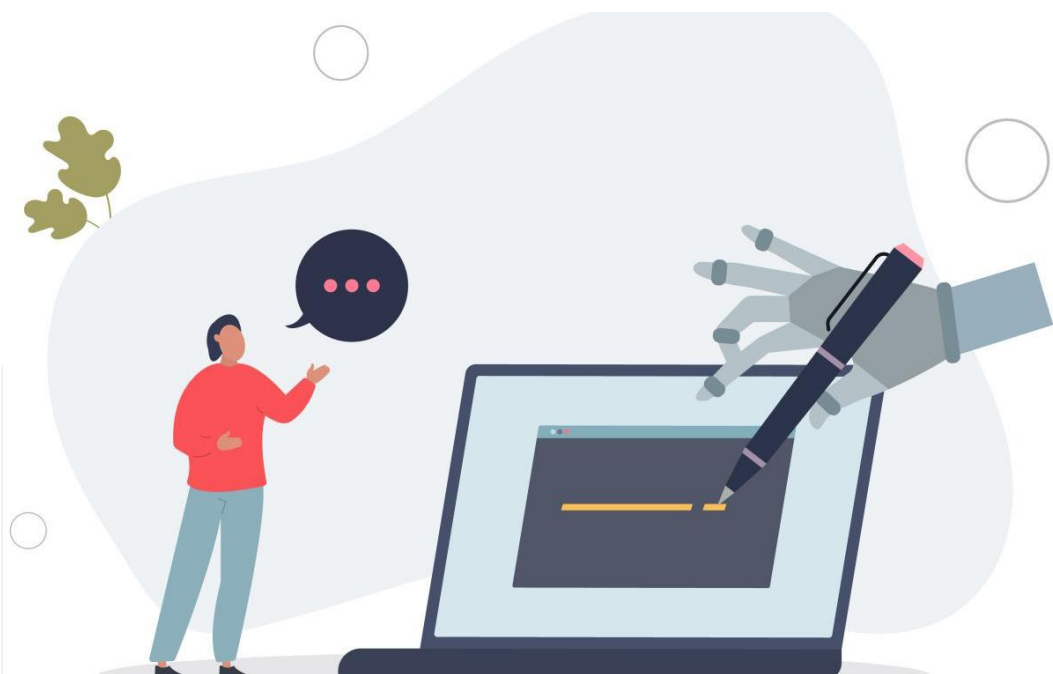


实验设计与文本智能筛选模块



实验设计

模块整合实验参数设置、文本分析与智能推荐功能，帮助研究者定义实验目标与变量。



文本智能筛选

利用AI分析候选文本特征，快速筛选最适合的实验材料，提升研究效率与准确性。

平台效果演示

译脑智图

翻译认知过程AI分析与可视化平台

首页 [文本分析](#) 数据处理 可视化

文本智能分析模块

本模块可帮助研究者分析候选文本的语言学特征，评估文本难度和复杂度，识别潜在的翻译难点，从而为翻译实验设计提供支持。

候选文本输入

文本标题

软件安装指南

文本类型

技术文档

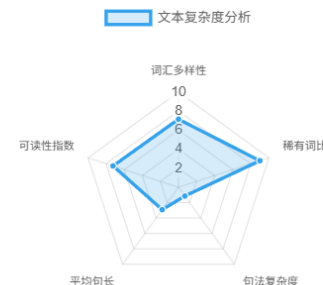
文本内容

To install the software, first ensure your system meets the minimum requirements: 2.5 GHz dual-core processor, 8 GB RAM, and 250 GB available hard disk space. Before installation, terminate all running applications and temporarily disable any antivirus software. Download the installation package from the official repository and verify the checksum to ensure file integrity. Run the installer with administrator privileges and follow the on-screen instructions. When prompted, select the "Custom Installation" option to configure installation directories and component selection. After installation completes, restart your system and run the configuration wizard to optimize performance settings based on your hardware profile.

分析文本特征

加载示例文本

文本复杂度评估



词频分布



多模态数据同步与清洗模块



多模态数据同步

自动同步不同设备记录的时间序列数据，解决数据整合难题，确保翻译过程研究的连贯性。

数据清洗

智能识别并标记异常数据点，生成标准化的数据结构，便于后续分析提升研究质量。

平台效果演示

译脑智图

翻译认知过程AI分析与可视化平台

[首页](#) [文本分析](#) [数据处理](#) [可视化](#)

多模态数据同步与清洗模块

本模块可帮助研究者同步和清洗眼动、键盘等多模态数据，自动识别异常数据点，并生成标准化的数据结构，为后续分析提供基础。

数据上传

眼动数据 (CSV)

[选择文件](#) 未选择任何文件

键盘日志数据 (CSV)

[选择文件](#) 未选择任何文件

[加载示例数据](#)

[重置](#)

处理参数设置

时间同步容差 (毫秒)

50ms 100ms 500ms

最小注视时长 (毫秒)

50ms 80ms 200ms

最大注视时长 (毫秒)

500ms 1000ms 2000ms

数据处理流程

1 数据加载与预览

加载眼动和键盘数据，识别关键列，展示基本信息

眼动数据

数据点数: 10
识别列: GazeX, GazeY, GazeEventType, GazeEventDuration
注视点数: 8
跳视点数: 2

键盘数据

数据点数: 18
识别列: timestamp, event_type, key, content
按键事件: 17
暂停事件: 1

2 数据清洗与异常检测

识别并移除眼动数据中的异常点，标准化数据格式

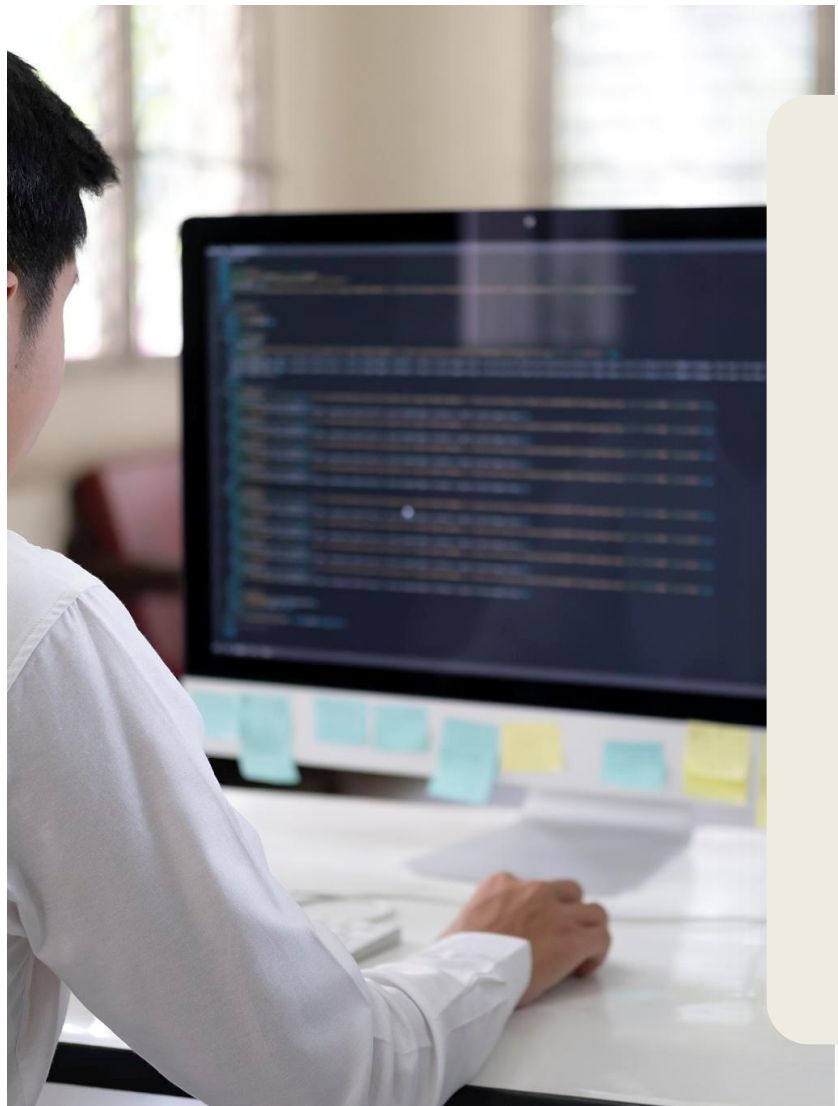
处理前数据点

总数据点: 10
注视点: 8
跳视点: 2

处理后数据点

有效数据点: 10
移除点数: 0
保留率: 100%

统计模型推荐与代码生成



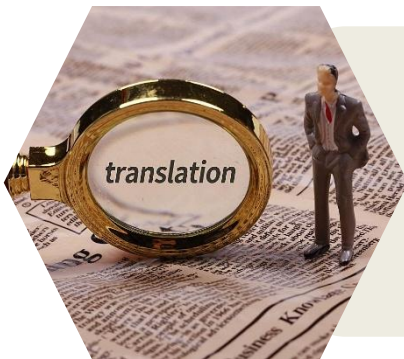
智能推荐

对于不熟悉统计分析的研究者，模块提供基于研究问题的分析方法智能推荐。

自动生成代码

模块能自动生成R或Python统计分析代码，预设分析模板，降低统计分析门槛。

认知过程可视化模块



翻译决策树

生成翻译决策树，展示译者决策路径，揭示翻译过程中的选择机制与策略运用。

认知负荷热图

创建认知负荷热图，显示文本难点分布，帮助译者识别并应对翻译中的认知挑战。



注意力流动图

构建注意力流动图，追踪译者视线轨迹，量化翻译过程中的注意力分配与移动模式。

平台效果演示

可视化示例

加载示例数据

示例1: 学术文本翻译

源文本

The development of neural machine translation has revolutionized the field of automated translation systems. Unlike traditional statistical models, neural networks can capture complex linguistic patterns and contextual information, leading to more fluent and accurate translations. However, these systems still face challenges with rare words, domain adaptation, and computational efficiency.

译文

神经机器翻译的发展彻底革新了自动翻译系统领域。与传统统计模型不同，神经网络能够捕捉复杂的语言模式和上下文信息，从而产生更流畅、更准确的翻译。然而，这些系统在处理罕见词汇、领域适应和计算效率方面仍然面临挑战。

翻译决策树

认知负荷热图

注意力流动图

决策树设置

决策类型显示

- ☒ 查询决策
- ☒ 修改决策
- ☒ 确认决策

时间轴

水平时间轴

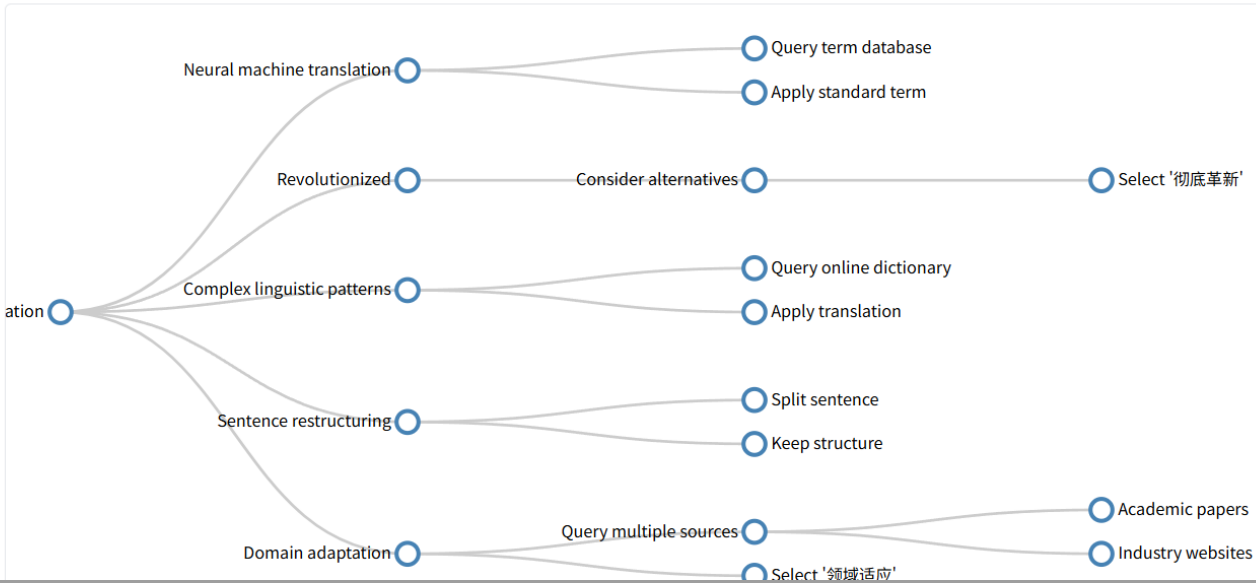
决策细节

Sentence restructuring

类型: 结构调整

时间: 40秒

调整句子结构以符合中文表达习惯



教学应用模块



认知模式分析

平台提供学习者认知模式分析工具，深入了解学习者在翻译过程中的认知特点与需求。



个性化教学建议

基于数据的个性化教学建议，为学习者提供定制化的翻译教学方案，提升教学效果。



教学案例库

平台建立翻译过程可视化教学案例库，涵盖不同文本类型与难度级别，丰富教学手段与资源。





5. 平台初步实现 与技术路线

技术栈选择

AI 集成

集成OpenAI API，提供智能分析与推荐功能，为平台注入AI力量，提升智能化水平。

后端

选用Python语言，负责数据处理与分析任务，确保平台数据处理高效准确。

数据可视化

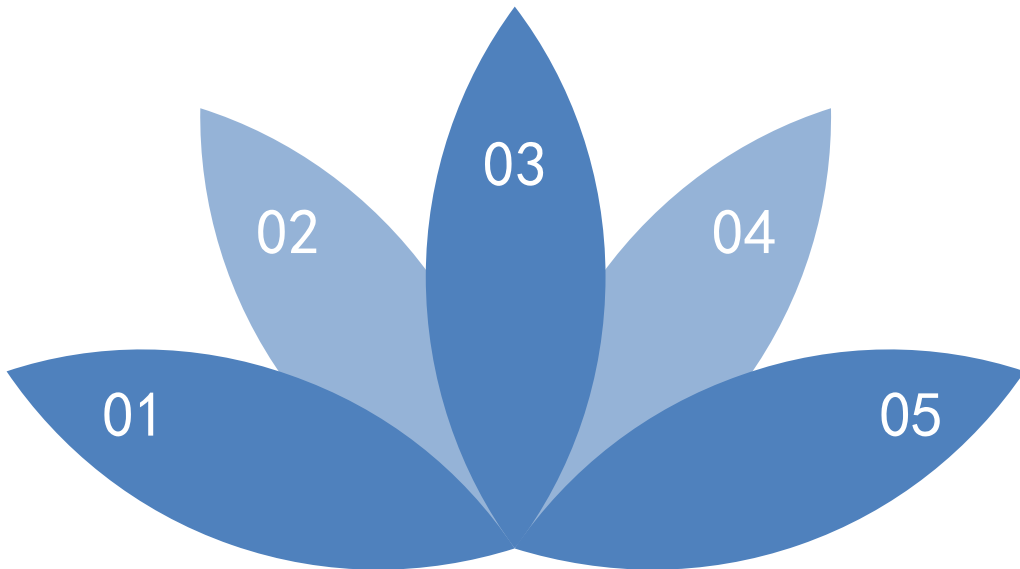
利用Plotly与Matplotlib生成交互式图表，让数据洞察直观易懂，助力决策。

前端

采用Streamlit框架，打造交互式Web界面，让研究数据可视化，提升用户体验。

部署

借助GitHub Pages与Streamlit Cloud，实现平台的便捷部署与分享，推动研究成果的应用。



开发路线与阶段目标



第一阶段（已完成）

完成核心架构的设计与技术选型任务，实现文本智能分析模块原型，并建立基础数据处理流程。



第二阶段（进行中）

优化多模态数据同步算法，实现统计模型推荐与代码生成功能，并初步集成可视化工具。



第三阶段（规划中）

开发认知过程的高级可视化功能及教学应用模块，同时优化用户界面并进行系统集成。

6. 小结



翻译认知过程可视化

翻译认知过程研究

翻译认知过程研究是探索译者思维活动、认知负荷和决策机制的重要领域。

翻译认知过程研究方法

采用眼动追踪、键盘记录、脑电图、屏幕录制和有声思维等方法收集数据。

翻译研究现状与挑战

翻译认知过程研究面临技术门槛高、数据处理复杂、分析工具缺乏和结果可比性差等挑战。



翻译认知过程可视化

新翻译认知研究平台

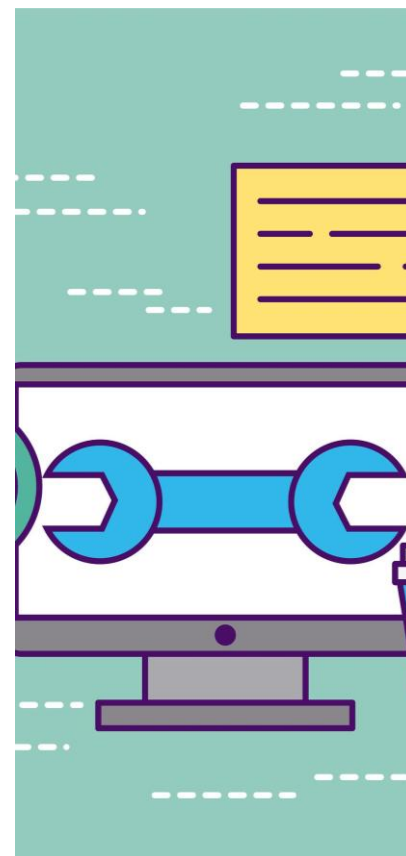
开发新的翻译认知过程研究平台可降低技术门槛、提高数据处理效率、拓展分析维度等。

AI辅助翻译研究

AI与翻译研究交叉创新，辅助实验设计、多模态数据处理和认知过程解析与可视化。

译脑智图平台

译脑智图平台采用模块化设计，包含实验设计、数据同步、统计模型、认知过程可视化与教学应用模块。



未来展望与研究价值

方法论创新

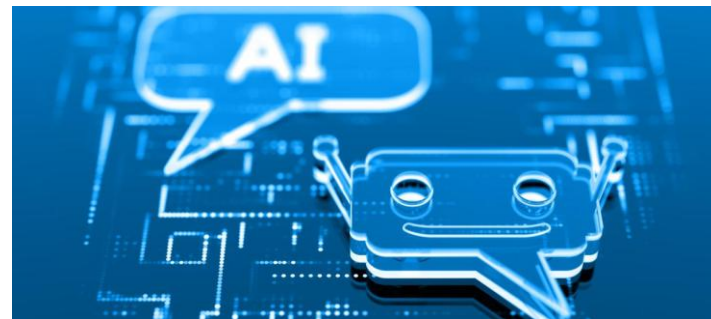
持续探索并开发混合研究方法，以更全面地揭示翻译认知的复杂机制，为翻译学研究提供多元化视角。

大规模数据分析

构建翻译认知过程数据库，支持跨语言、跨群体的比较研究，推动翻译认知科学的进步。

教学应用拓展

将平台整合到翻译硕士培养中，形成技术学习、研究实践、成果产出的闭环，促进翻译教育的创新发展。



THANKS

黄婕

对外经济贸易大学

<https://chezvivian.github.io>