**基于langChain 构建智能体Agent**

前段时间，随着 [GPT-4o](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=GPT-4o&zhida_source=entity)、Sora 的陆续问世，[多模态模型](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%A4%9A%E6%A8%A1%E6%80%81%E6%A8%A1%E5%9E%8B&zhida_source=entity)在生成式方面取得的成绩无可否认，而人工智能的下一个革命性突破将从何处涌现，引起了大量学者和相关人士的关注。

人工智能著名学者、斯坦福大学教授[吴恩达](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%90%B4%E6%81%A9%E8%BE%BE&zhida_source=entity)一直非常推崇智能体。此前他曾在个人[博客](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%8D%9A%E5%AE%A2&zhida_source=entity)着重指出「AI 智能体工作流将会在今年推动人工智能取得长足进步」，AI 智能体的未来潜力愈加被看好，吸睛无数。

近日，吴恩达延续他之前的脚步，[开源](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%BC%80%E6%BA%90&zhida_source=entity)了一个 AI 智能体机器翻译项目。



项目链接：

<https://github.com/andrewyng/translation-agent>

人工智能专家吴恩达开源了一个全新的 AI 智能体机器翻译项目Translation Agent。据介绍，Translation Agent 是一款基于反射式代理工作流程的机器翻译工具，利用 LLM 技术提供高度定制化的翻译服务。

这个项目具有高度可控的翻译工作流，用户只需简单调整提示词，就能灵活设定语气、地区特色，并可导入专业术语表以确保翻译术语的一致性。这种灵活性和可定制性，将使 AI 智能体在机器翻译领域发挥更加广泛的应用。

**主要功能点:**

提供反射式代理工作流程，通过 LLM 进行翻译，反思并提出改进建议，从而改进翻译质量。

可定制性强，允许根据需要修改输出风格、处理特殊术语和指定特定区域或方言的语言使用。

适用于不同语言对的翻译，通过调整提示和选择不同的 LLM，可以实现更优质的翻译效果。

**使用场景 :**

企业通信:为企业提供准确、定制化的翻译服务，助力跨语言沟通和国际合作。

学术研究:支持跨语言学术交流，帮助研究人员和学者分享知识和合作成果。

个人生活:为个人提供便捷、准确的翻译工具，解决语言障碍问题，促进文化交流。

**工作流程 :**

提供源语言和目标语言信息以及要翻译的文本。

使用 LLM 进行翻译，并生成初步翻译结果。

LLM 反思翻译，提出改进建议。

他分享了关于 AI 智能体机器翻译对改进传统[神经机器翻译](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E7%A5%9E%E7%BB%8F%E6%9C%BA%E5%99%A8%E7%BF%BB%E8%AF%91&zhida_source=entity)方面的看法：「具有巨大潜力，尚未被完全发掘」，并发布了一个他一直在周末玩的翻译智能体演示。该翻译智能体以 MIT 许可证形式发布。用户可以自由使用、修改和分发该代码，无论是商业用途还是非商业用途。

在研究团队有限的测试中，吴恩达团队开源的翻译智能体有时能够与领先的商业提供商进行同等水平的竞争，有时则不如它们。但它仍提供了一个高度可控的翻译系统，只需简单更改 [prompt](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=prompt&zhida_source=entity)，使用者就可以指定语气（正式 / 非正式）、地区变体（例如：使用者想要在西班牙本地说的西班牙语，还是在拉丁美洲说的），并确保术语的翻译一致性（通过提供[词汇表](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E8%AF%8D%E6%B1%87%E8%A1%A8&zhida_source=entity)）。这个应用程序虽然目前仍稍显稚嫩，但鉴于反思工作流已经展现出不错的成果，吴恩达认为智能体翻译仍有很大的提升空间。

在吴恩达分享的[开源项目](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%BC%80%E6%BA%90%E9%A1%B9%E7%9B%AE&zhida_source=entity)中，具体介绍了 AI 智能体翻译项目的工作流。

**翻译智能体：使用反思工作流进行智能体翻译**

这是一个使用 Python 演示的反思智能体工作流的机器翻译示例。主要步骤如下：

1. 输入 prompt，使[大型语言模型](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%A4%A7%E5%9E%8B%E8%AF%AD%E8%A8%80%E6%A8%A1%E5%9E%8B&zhida_source=entity)（[LLM](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=LLM&zhida_source=entity)）将文本从 source\_language 翻译成 target\_language；

2. 让 LLM 反思翻译结果，并提出建设性的改进建议；

3. 利用这些建议改进翻译。

**自定义能力**

通过使用 LLM 作为[翻译引擎](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E7%BF%BB%E8%AF%91%E5%BC%95%E6%93%8E&zhida_source=entity)的核心，该系统具有高度可控性。

例如：通过更改 prompt，这种工作流比传统的[机器翻译](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=5&q=%E6%9C%BA%E5%99%A8%E7%BF%BB%E8%AF%91&zhida_source=entity)（MT）系统更容易实现以下功能：

* 修改输出的风格，如正式 / 非正式。
* 指定如何处理习语和特殊术语，如名字、技术术语和缩写。例如，在 prompt 中包含术语表，可以确保特定术语（如开源、H100 或 GPU）翻译的一致性。
* 指定特定区域的语言使用或特定方言，以服务目标受众。例如，拉丁美洲的西班牙语与西班牙的西班牙语不同；加拿大的法语与法国的法语不同。

通过 [BLEU](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=BLEU&zhida_source=entity)（Bilingual Evaluation Understudy）分数进行翻译质量的评估作为有别于传统机器翻译的后起之秀，同样也是衡量 AI 智能体翻译的必由之路。

根据使用传统翻译数据集的 BLEU 分数进行评估的结果：这种工作流有时能与领先的商业产品竞争，但有时表现也不如它们。不过，它偶尔也能够得到非常好的结果，甚至优于商业产品。

他们认为这只是智能体翻译的起点，这一方向在翻译方面很有前景，并且有很大的改进空间。因此，作者团队发布这一演示，以鼓励更多的讨论、实验、研究和开源贡献。

如果相比于更快且更便宜的传统架构（例如输入文本并直接输出翻译的端到端 Transformer 架构），智能体翻译能够有更好的结果，那么它就 neng 提供一种自动生成训练数据（[平行文本](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%B9%B3%E8%A1%8C%E6%96%87%E6%9C%AC&zhida_source=entity)语料库）的方法，可以用于进一步训练和改进传统算法。

**翻译智能体的进一步发展**

吴恩达在最后还分享了几点希望[开源社区](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%BC%80%E6%BA%90%E7%A4%BE%E5%8C%BA&zhida_source=entity)能够尝试的想法，希望凭借众家之力将翻译智能体的巨大潜力归为实处。

* 尝试其他语言[生成模型](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E7%94%9F%E6%88%90%E6%A8%A1%E5%9E%8B&zhida_source=entity)。此项目主要使用 [gpt-4-turbo](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=gpt-4-turbo&zhida_source=entity) 进行原型开发。其他人可以尝试其他 LLM，以及其他[超参数](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E8%B6%85%E5%8F%82%E6%95%B0&zhida_source=entity)选择，并查看是否有些大模型可以对特定语言更好地翻译。
* 术语表的创建。使用 LLM 也许可以更高效地建立术语表。例如，许多企业使用的是互联网上不常用的专业术语，而 LLM 可能不知道这些术语。此外，还有许多术语可能有多种翻译方式。例如，「open source」在西班牙语中可以是「Código abierto」或「Fuente abierta」；两者都可以，但最好选择一个并在单个文档中坚持长期使用。
* 术语表的使用和实施。将术语表包含在 prompt 中最好的方式是什么？
* 在不同语言上进行评估。翻译智能体在不同语言中的表现会发生怎样的变化？有没有通过一些变动，使其在特定源语言或目标语言上表现更好的方法？（请注意，对于 MT 系统正在接近的较高性能水平，BLEU 是否是一个很好的度量标准仍是不确定的。）此外，对于资源较少的语言，它的性能表现仍需要进一步研究。
* [错误分析](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E9%94%99%E8%AF%AF%E5%88%86%E6%9E%90&zhida_source=entity)。吴恩达团队发现此应用程序对于一些指定语言和国家 / 地区（例如，「在墨西哥作为普通话的西班牙语」）来说效果很好。除此之外，当前方法在哪些方面仍存有不足？翻译智能体在专业主题（如法律、医学）或特殊文本类型（如电影字幕）上的性能表现如何？存在怎样的限制？
* 更好的评估指标。吴恩达认为对 AI 智能体翻译进行更好的评估是一个巨大且重要的研究课题。与其他生成自由文本的 LLM 应用程序一样，当前的评估指标似乎并不足够评估翻译智能体的表现。例如，他们发现：即使在主动型工作流程在捕捉上下文和术语方面表现更好的文档上，仍会导致人类评分者更喜欢当前的商业产品，但是在句子级别进行评估（使用 FLORES 数据集）时，主动型系统的 BLEU 得分则较低。在设计出更好的度量标准（也许使用 LLM 评估翻译？）以在文档水平上更好地实现与人类偏好相关的翻译质量仍需更多的努力。

值得注意的是，一些学术研究小组也开始关注基于 LLM 和主动型翻译的研究。

对于 AI 翻译智能体的前景，吴恩达认为这个领域还处于起步阶段，并分享了一些相关的学术论文供大家参考。

对于 AI 翻译智能体的前景，吴恩达认为这个领域还处于起步阶段，并分享了一些相关的学术论文供大家参考。

* 论文标题：ChatGPT MT: Competitive for High- (but not Low-) Resource Languages
* 论文地址：[https://arxiv.org/pdf/2309.07423](https://link.zhihu.com/?target=https%3A//arxiv.org/pdf/2309.07423)
* 论文标题：How to Design Translation Prompts for ChatGPT: An Empirical Study
* 论文地址：[https://arxiv.org/pdf/2304.02182v2](https://link.zhihu.com/?target=https%3A//arxiv.org/pdf/2304.02182v2)
* 论文标题：Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts
* 论文地址：[https://arxiv.org/pdf/2405.11804](https://link.zhihu.com/?target=https%3A//arxiv.org/pdf/2405.11804)

吴恩达此次开源的翻译智能体仍处于初级阶段，但已在机器翻译数据集上有了较好的表现，为 [AI 智能体](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=8&q=AI+%E6%99%BA%E8%83%BD%E4%BD%93&zhida_source=entity)的下一步发展注入了又一剂[强心针](https://zhida.zhihu.com/search?content_id=244391527&content_type=Article&match_order=1&q=%E5%BC%BA%E5%BF%83%E9%92%88&zhida_source=entity)。

相关信息参考

**字幕翻译、书籍翻译的福音，如何借助 ChatGPT 得到高质量的翻译结果？**

<https://baoyu.io/blog/prompt-engineering/how-to-get-high-quality-translation-results-with-chatgpt?continueFlag=a7465afbb79571404de2ee86b4f6409b>

**什么时候该用多智能体是不是一定要用多智能体？**

<https://baoyu.io/blog/ai/when-to-use-multi-agent-systems-or-cot?continueFlag=a7465afbb79571404de2ee86b4f6409b>

**改变传统，吴恩达开源了一个机器翻译智能体项目**

<https://zhuanlan.zhihu.com/p/703174520>

作业要求：

1: 阅读吴恩达Agent 源代码 ，部署运行 **（必需）**, 可以增加部分功能（**可选项**）

A： 可以考虑多模态支持，上传音频文件进行利用大模型转换文本文件，然后智能体翻译

B：可以支持语音录制，然后翻译

C：上传带有需要翻译的 图片（包含需要翻译的文本）

2：参考吴恩达实现原理， 自己重头构建一个全新的智能翻译（**可以仅仅实现部分功能**）