

CONE多语言翻译

CONE是一个中文为核心的轻量级多语言翻译模型，后续会将模型开源。CONE模型是在NLLB-1.3B的基础上，由300+语言数据训练得到的单个翻译模型。我们会逐步放出所有语言进行公开测评。如果您觉得翻译有误或者有更多的翻译建议，可以在翻译页面进行反馈，我们非常期待您的意见！

公测进展

- 2023年3月18日，50个语种通过内部测试进入公开测评。

目录

- 导言
- 模型亮点
- 模型缺点
- 相关产品比较
- CONE性能
- 对比开源翻译模型
- 对比商业翻译软件
- 样例展示
- 翻车样例展示
- 常见问题及解答

导言

语言是人类交流和表达思想的重要工具，不同的语言也反映了不同文化之间的差异。从遥远的非洲大陆到到寒冷干燥的北欧，从湿润的地中海到充满异域风情的中东内陆，从遍地矿产的俄罗斯到物产富饶的东亚，从古文明古国到热情奔放的南美大陆，来自不同地区的人们每天使用着不同的语言进行生活。据统计，全球共有约7,000种语言。随着全球化的进程和人类文明的发展，跨语言的交流变得尤为重要。

CONE致力于推动支持更多语言的翻译。无论是开源模型还是商业翻译软件，目前覆盖的语言种类仍然十分有限。实际上，许多语言甚至没有任何可用的机器翻译系统。对于那些使用非常少见语言的人来说，获得高质量的机器翻译系统就变得非常重要。例如，少数民族地区的人们可能会使用一种特定的语

言，但是他们需要与其他地区的人进行交流。翻译技术不仅有助于促进商业往来，还能够促进文化交流和理解。但是每种语言在语法、词汇和表达方式上差异较大，要让计算机能够准确地翻译这些语言，需要大量的数据、人工标注、计算资源等。因为小语种的使用者数量较少，存在数据不足，人工标注缺乏等问题，CONE也在致力于提升小语种的翻译数量和质量。

模型亮点

1. 开源。CONE模型会开源，用户可以根据自己的需求定制化翻译工具，也可以根据翻译内容开发多语言相关工具。
2. 与现有开源模型和商业系统相比，CONE支持的语言数目领先。
3. 中文为核心的翻译性能进行了额外优化。

模型缺点

1. 小语种翻译性能仍然是较大挑战。根据人工评测结果，中文为核心的翻译结果均分在3分以上（满分5分）的语言仅有186个。很多语言仍然面临翻译不准确的问题。
2. 暂不支持古诗词翻译、文言文翻译。
3. 不同于商业模型，人名地名组织名暂时没有进行额外处理。

相关产品比较

为了更好的理解和量化CONE的能力，以下是对相关产品的一些介绍：

1. chatGPT类语言模型：

优点：根据现有研究结果和我们的实验结果，chatGPT在德英大语向上性能可以逼近领先监督翻译模型效果，小语种时不时也有惊艳的翻译效果。

缺点：支持语种数量比较有限；模型参数多，推理代价大；小语种总体支持相对较弱。

2. 商业翻译软件：

优点：核心语言积累较多，用户体验好。

缺点：大部分的商业软件仅支持几十种或者上百种语言。百度是唯一支持200种语言的翻译产品；大规模使用有经济成本。

3. 开源翻译模型：

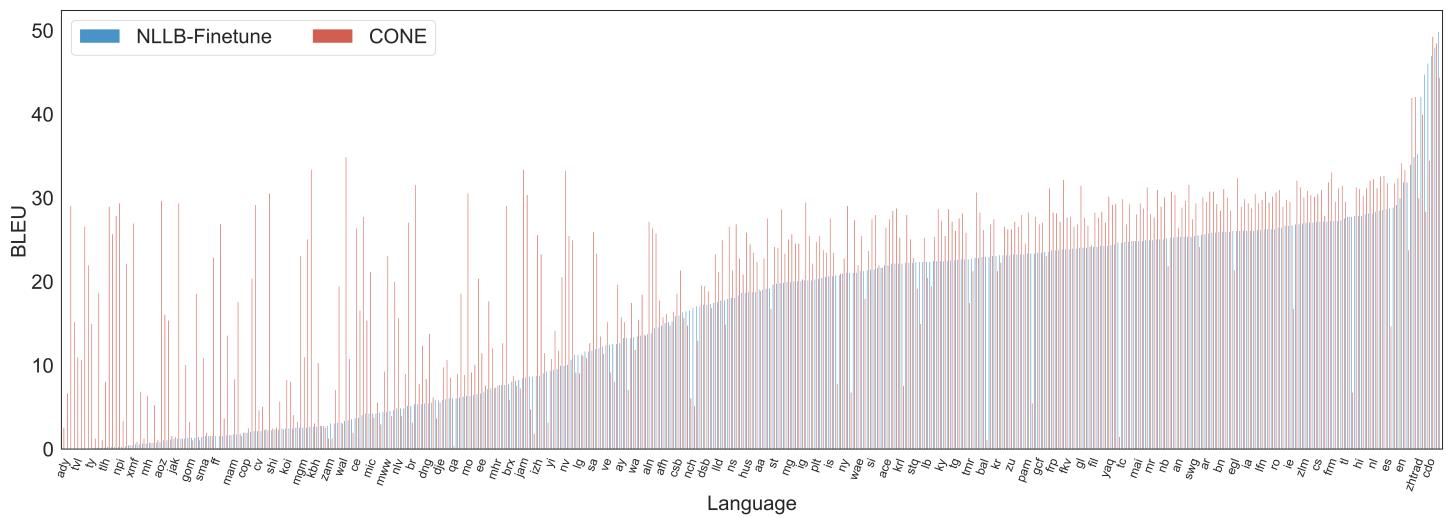
优点：免费；可定制化；后续研究可以复用开源工作。

缺点：翻译性能相对比较有限；比如领先开源机器翻译模型NLLB对非拉丁语系支持有限，中文翻译性能较差；性能较好的翻译模型比如NLLB-54B推理代价较大。

以上是一些定性的介绍，具体的性能比较可以参考下文中模型对比部分。机器翻译是自然语言处理中积累比较多的垂直领域，chatGPT目前也仅只能在有限的语向上达到领先翻译结果。但是由于GPT类模型是通用模型，领先的翻译效果也展示了通用模型用于垂直领域的潜力，下一步我们将会结合机器翻译和开源的通用模型译进行垂直领域新探索。

CONE性能

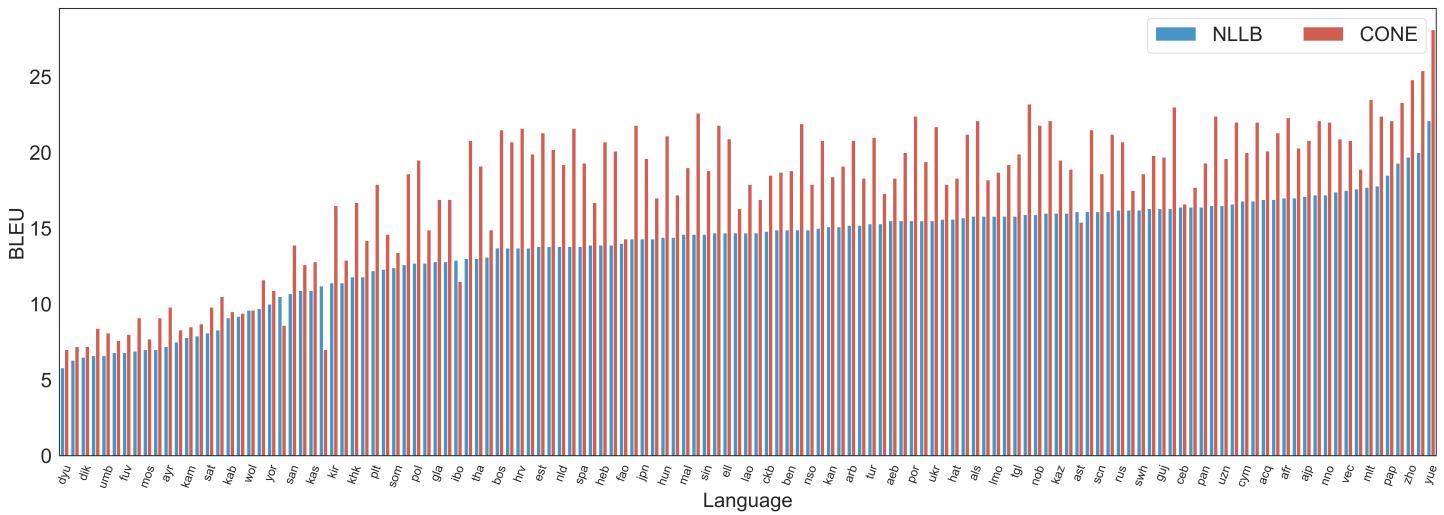
小语种的评测一直是比较大的难题，我们在中文数据（FLORES）上利用回译的方式（根据翻译到目标语言再翻译回原语言）计算原文和回译内容BLEU。我们和NLLB-Finetune（NLLB-1.3B模型直接在我们语料上训练得到的模型）进行了比较，以下是评测结果：



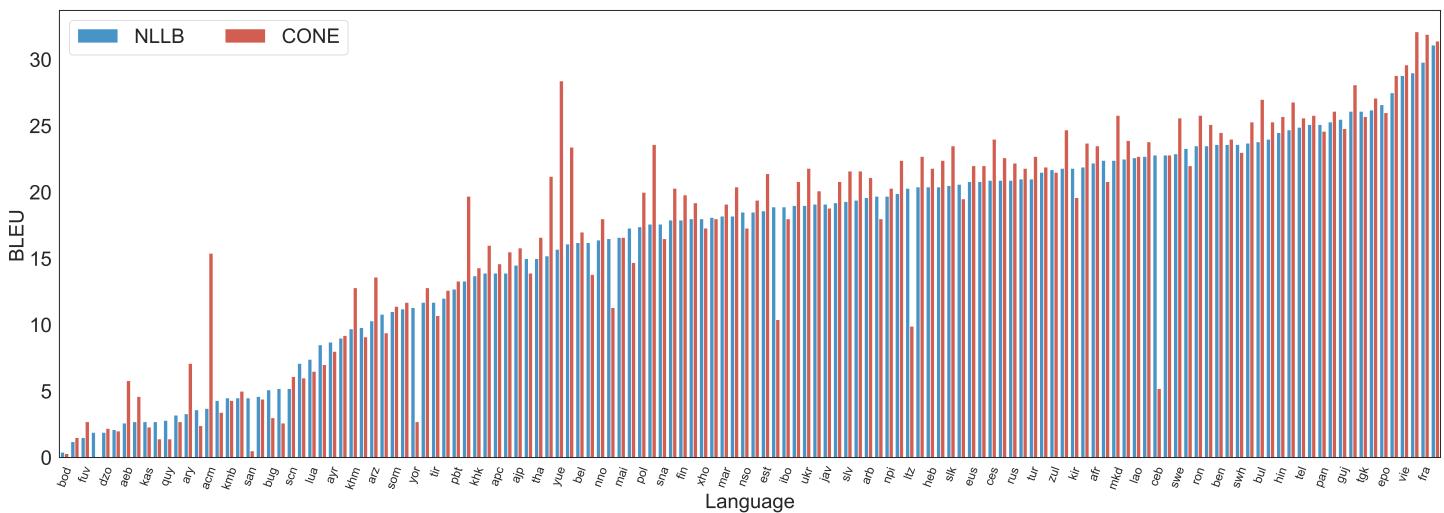
对比开源翻译模型

CONE模型是在NLLB-1.3B的基础上在包含300+语言的数据上训练得到的新模型。这里和NLLB的比较并不是公平的比较，只是想对比CONE模型在支持更多语言的同时保有竞争力的翻译性能。我们在flores-200上进行了评测，评测数据为CONE训练数据和NLLB训练数据重合的139种语言。

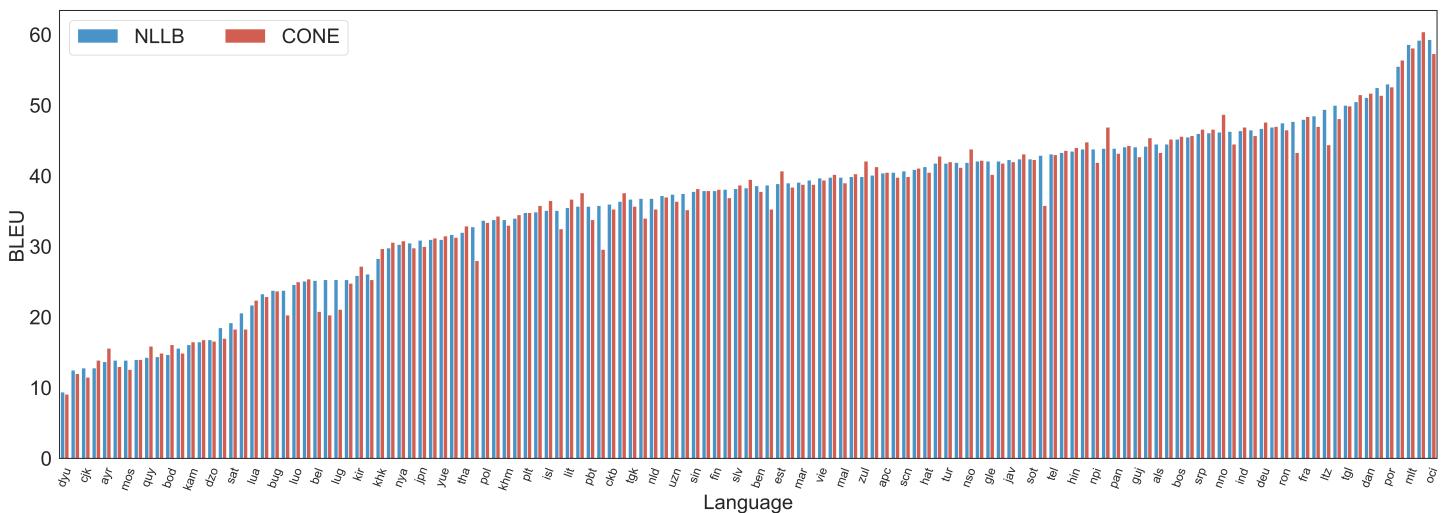
其他语言到中文翻译效果：



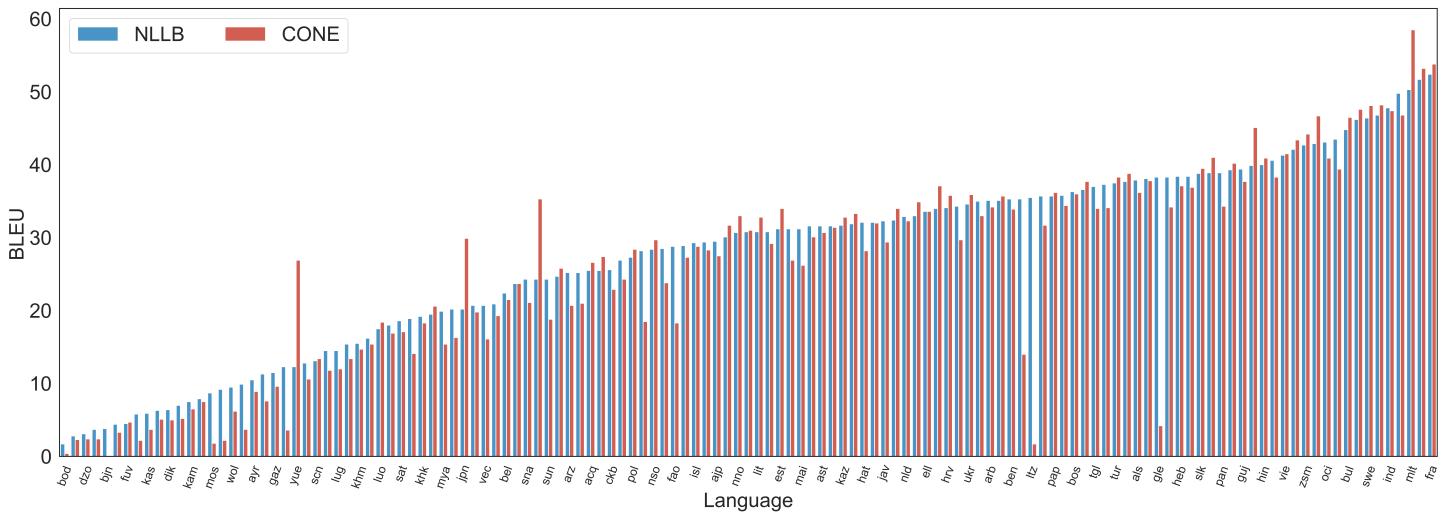
中文到其他语言翻译效果：



其他语言到英语翻译效果：



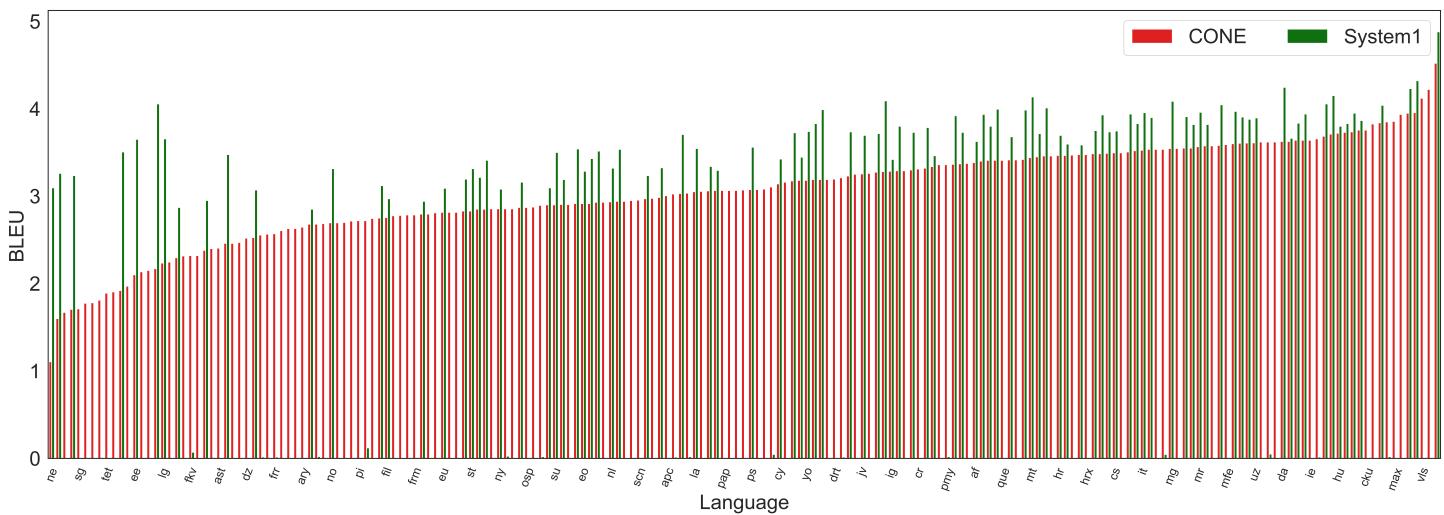
英语到其他语言翻译效果：



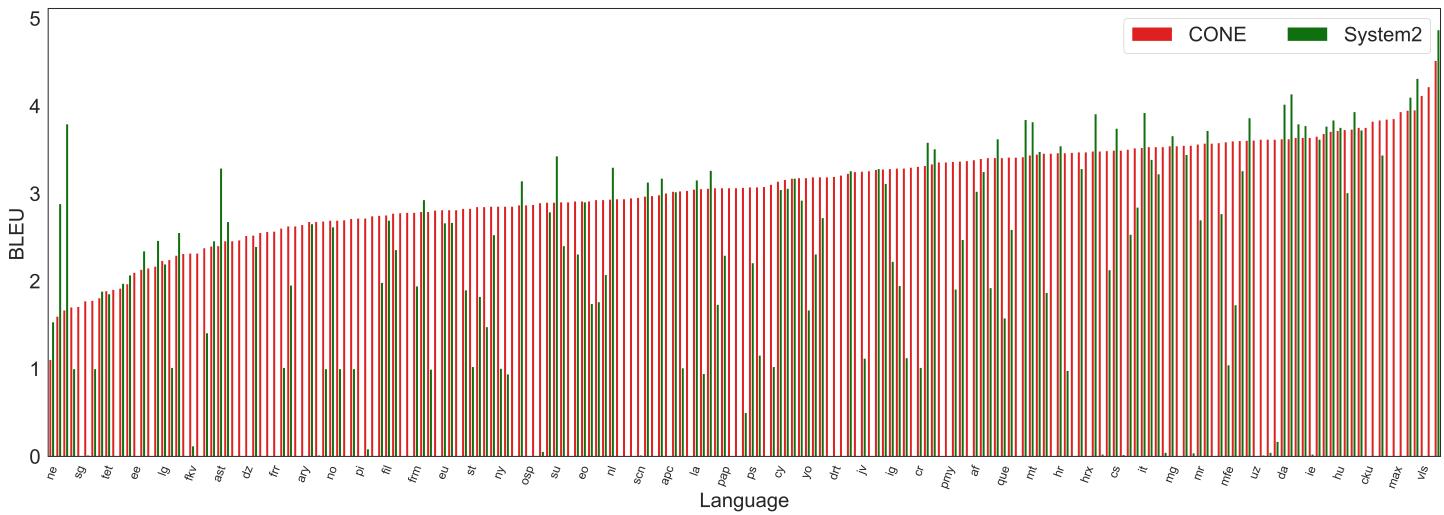
对比商业翻译软件

我们在中文为核心的翻译上选择了国外知名商业翻译软件(A)和国内知名商业翻译软件(B)进行比较。为了更准确地衡量翻译效果，我们利用了中文评测人员优势，选择100条中文数据进行回译评测。我们选择了商业系统A、B、CONE和NLLB 1.3B的共同覆盖语言。回译评测指通过将中文翻译成对应语言，再将对应语言翻译回中文，通过对比中文语义匹配程度进行比较。每个样本会同时分发给两位不同的标注员。评测分数包括0-5分。0分表示不支持的语言或翻译内容完全不能接受，1分表示仅能匹配很少一部分单词，2分表示仅有少部分语义可以匹配，3分表示大部分语义可以匹配，4分语义基本还原，5分表示不仅语义正确，表达也非常流畅。**在中文为核心的翻译中，商业系统A平均好于CONE，CONE平均好于商业系统B。**

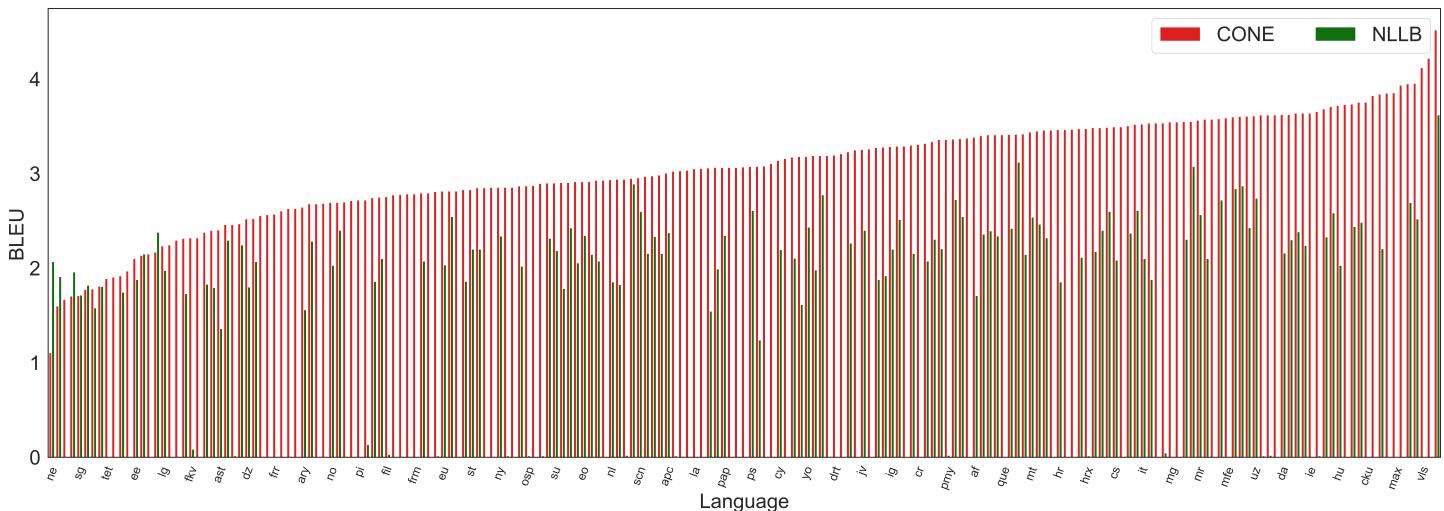
和商业系统A的比较结果：



和商业系统B的比较结果：



除此之外，我们还比较了起始模型NLLB-1.3B的翻译效果。由于NLLB-1.3B为起始模型，这里并不是公平比较，只是为了说明开源系统和商用系统在中文核心语向上存在的差距。以下是和NLLB的比较结果：



商业系统背后所用的技术、数据处理流程更加复杂。CONE仅使用单个模型也可以在部分语向达到商用翻译系统性能。

样例展示

此处展示了一些CONE比较擅长的翻译案例：

翻译领域	输入文本	CONE翻译结果
日常简单对话	大郎，该吃药了	Dude, you need to take the pill.
歌曲翻译	我真的还想再活500年	I really want to live another 500 years.

翻译领域	输入文本	CONE翻译结果
电视剧台词 《觉醒时代》	生逢乱世，即使命运如蝼蚁， 但仍有人心向光明。	The world is a mess, but the light is still there.
电视剧台词 《觉醒时代》	他们可以挡住天上的太阳， 但他们无法挡住民主的光芒。	They can block the sun in the sky, but they cannot block the light of democracy.
电视剧台词 《阿甘正传》	Mama always said life was like a box of chocolates. You never know what you're gonna get.	妈妈总是说生活就像一盒 巧克力,你永远不知道你会得到什么。
百科	人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。	Artificial intelligence is a branch of computer science that seeks to understand the essence of intelligence and produce a new kind of intelligent machine that can react in a similar way to human intelligence, and the field includes robotics, language recognition, image recognition, natural language processing and expert systems.
百科	ChatGPT是美国人工智能研究实验室OpenAI新推出的一种人工智能技术驱动的自然语言处理工具，使用了Transformer神经网络架构。	Chat GPT is a newly launched artificial intelligence-driven natural language processing tool from the US AI Research Laboratory OpenAI, using the Transformer neural network architecture.

此处展示了一些CONE比较不擅长的领域翻译案例：

领域	输入	翻译
古诗词	白日依山尽， 黄河入海流	The day of the flood, the day of the flood
古诗词	空悲切，白了少年头	The white, the youthful.

领域	输入	翻译
专有名词 (人民地名组织名)	鲁迅是个帅小伙	Rusty is a good boy.
结构复杂	我一把把把把住了	I put a stop to it.
领域名词	溜肥肠溜肝尖	Flat-Hearted

常见问题及解答

Q1：内测服务为什么会返回输入错误和生成内容错误？

为了避免翻译系统被用来翻译不当内容，我们调用了阿里云的敏感词过滤接口。如果出现误触发问题，请联系cone@x-venture.tech进行反馈。

Q2：为什么模型在小语种的翻译质量会出现翻车案例？

尽管我们优化了小语种的翻译能力，但是小语种的翻译性能和大语种的翻译性能还是有明显差距。如果您觉得翻译效果不如预期，可以在翻译页面选择喜欢和不喜欢，也可以联系cone@x-venture.tech进行反馈。

Q3：为什么对外服务还有预约？

由于内测服务资源有限，我们只能在同一时间段对有限的用户提供服务，敬请谅解。

Q4：我是专业的小语种用户，发现问题怎么反馈？

如果您是小语种从业者，欢迎联系cone@x-venture.tech。我们也正在积极招募小语种标注员和运营实习生。

Q5: CONE的创新点是什么？

CONE项目独立于研究工作，主要是提供轻量级翻译开源工具。

Q5: CONE不擅长什么？

CONE的训练语料暂不支持古诗词、文言文的翻译；对一些复杂文本结构：“我一把把把住了”，“干一行行行行”支持比较有限；没有对人名地名做额外优化。