

QCon
全球软件开发大会

Hunyuan-MT: 腾讯混元轻量级大模型的翻译训练 与优化实践

郑茂

腾讯混元技术专家，混元应用算法负责人



自我介绍



郑茂，现任腾讯混元技术专家，混元应用算法负责人。负责混元大模型在腾讯的应用落地，包括广告、社交、娱乐、教育、客服等业务的大模型应用层效果优化。

技术研究方向包括大模型复杂推理、多轮对话、翻译、RAG、AGENT 等。在 ACL、AAAI、EMNLP、COLING、CVPR 等会议上发表数十篇论文。带领团队获得 WMT2025 翻译比赛诸多赛道第一名。

目录

01

Hunyuan-MT近期进展

02

Hunyuan-MT CASE分享

03

Hunyuan-MT模型训练方案详解

04

Hunyuan-MT应用场景能力建设方案

05

经验总结

06

未来展望

📍 北京

QCon

全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

📍 上海

QCon

全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月

5月

6月

8月

10月

12月

📍 上海

AiCon

全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

📍 深圳

AiCon

全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOps
- 具身智能

极客邦科技 2025 年会议规划

促进软件开发及相关领域知识与创新的传播



参会咨询



查看会议

01

Hunyuanyuan-MT近期进展

关键进展

WMT2025

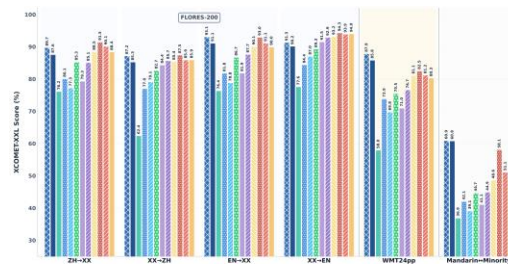
- 通用翻译赛道上，**31个语种中30个语种第一**（超过Gemini-2.5-pro、CLAUDE4、deepseek v3、qwen3 235b等模型）。
- 31个语种除了中文、英语、德语、日语等常见语种，也包含捷克语、马拉地语、爱沙尼亚语、冰岛语等小语种

English-Simplified Chinese									
System Name	LP Supported	Params (B)	Humeval?	AutoRank	CometKw	GEMBA-XL ↑	GEMBA-ESA-CMDA ↑	GEMBA-ESA-GPT4.1 ↑	XCOME MetricX-24-Hybrid-XL ↑
Shy-hunyuan-MT	✓	7	✓	1.0	0.67	87.2	88.3	-4.0	0.576
Wenyil	✓	14	✓	3.9	0.663	84.2	87.7	-5.0	0.52
▲ Gemini-2.5-Pro	✓	?	✓	4.0	0.657	85.2	88.7	-4.9	0.512
Alphab	✓	14	✓	4.1	0.66	84.7	87.8	-5.0	0.515
▲ GPT-4.1	✓	?	✓	4.6	0.652	84.9	86.8	-5.0	0.512
▲ Qwen3-235B	✓	235	✓	4.8	0.661	82.7	85.0	-5.0	0.513
Yulu	✓	14	✓	4.8	0.687	74.9	77.1	-4.6	0.542
GemTrans	✓	27	✓	4.9	0.658	77.0	80.2	-4.3	0.546
▲ Mistral-Medium	?	?	✓	4.9	0.658	82.4	84.9	-5.0	0.514
CommandA-WMT	✓	111	✓	5.6	0.665	78.9	81.5	-5.0	0.508
UvA-MT	✓	12	✓	6.3	0.671	76.8	81.0	-5.4	0.499
▲ Claude-4	✓	?	✓	7.0	0.649	80.4	82.8	-5.6	0.487
▲ DeepSeek-V3	✓	671	✓	7.1	0.658	84.9	85.1	-5.2	0.473
▲ Llama-4-Maverick	✓	400	✓	8.0	0.65	74.9	79.4	-5.5	0.489
▲ ONLINE-B	✓	?	✓	8.2	0.656	73.0	74.7	-5.2	0.492
▲ Gemma-3-27B	✓	27	✓	9.0	0.638	75.5	78.7	-5.8	0.475
Lanqo	✓	9	✓	9.1	0.665	63.6	67.7	-4.9	0.513

夺冠

Hunyuan-MT开源

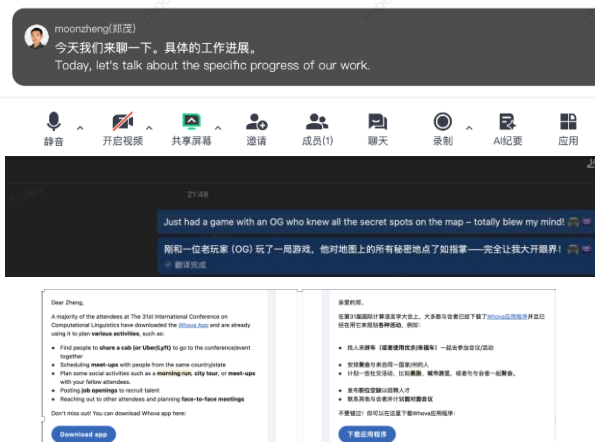
- 开源Hunyuan-MT-7b翻译模型和Hunyuan-MT-Chimera翻译集成模型。
- 重点支持33个语种互译。特色能力包括5种民汉语言/方言互译、俚语翻译等
- Huggingface首页趋势榜**排名第一**，开源一周模型下载量破万，开源一周Github Star数量超500



开源

多场景应用落地

- 接入腾讯会议（实时翻译、会后翻译），腾讯游戏出海翻译、企业微信（聊天记录翻译、邮件翻译），微信读书（划词翻译、书籍翻译）、QQ浏览器（文件翻译、划词翻译），腾讯海外客服、腾讯云翻译等。



应用

02

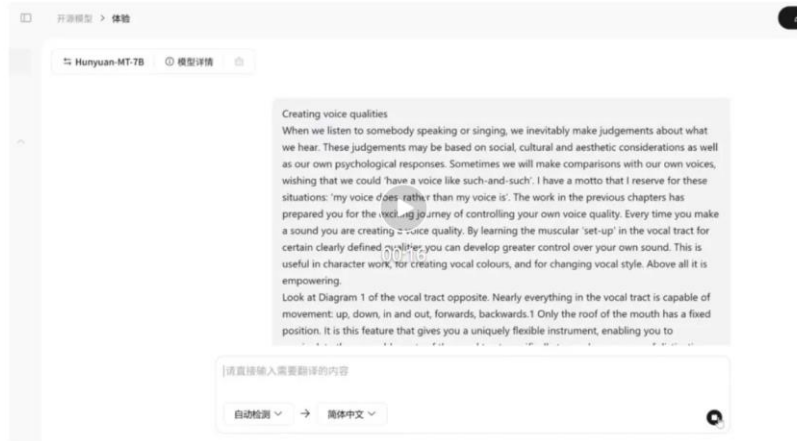
Hunyuan-MT CASE分享

Hunyuan-MT-7B

实测第一关

专业声乐领域的文献翻译，由于艺术的抽象加上专业术语，翻译难度具有代表性。

我给了它一大段文献里的内容。



速度非常快，翻译结果很准确，语言也不生硬。

当大模型翻译越来越没有人味，腾讯混元闷声干了件大事儿

原创 Bubble BubbleBrain 2025年09月07日 11:54 上海



翻译这件事儿，自从有了通用的大语言模型之后似乎就变得没那么重要了。

毕竟对大部分人来说，以GPT、DeepSeek、Gemini 之类的能力随手翻译个文章，报告什么的，真的信手拈来，大差不差看着也就行了。

但，腾讯不这么认为。

翻译界王炸模型出现，被字幕坑过的我终于等到救星！

原创 Loki2046 赛博小熊猫Loki 2025年09月08日 09:00 北京

本文字数3000+，图片51张，阅读时间约8分钟。

不知道你平常看不看英美日韩泰剧？

我在不忙的时候是很爱看的，有些节奏快、有些剧情曲折、有些日常温馨、有的狗血但就是吸引你往下看。

可以说不同的剧各有各的优点，但他们有一个共同的缺点——我听不懂！（好吧，这是我的缺点.....）

太好了，有救了！我的翻译搭子拿了全球第一



纯纯一个囫圇吞枣，你就感觉自己大概看懂了，但所有的词句都得自己重新斟酌、排序、组装。

我一直觉得机翻就这水平了，差不多看懂就行，还要什么自行车。

直到今天，实测了下Hunyuan-MT-7B。

不吹不黑，直接上案例。

1、长难句无压力

定语很多的长难句，混元毫无压力。

原文（摘自路透社）：

Trump's Republican allies and members of the administration, including Vice President JD Vance, took to social media to criticize lawmakers who had grilled the health secretary, "You're full of shit and everyone knows it," Vance said on X.

页面翻译：特朗普的共和党盟友和政府成员，包括副总统 JD Vance，在社交媒体上批评了抨击这位卫生部长的立法者，“你满身都是狗屎，每个人都知道，”万斯在 X 上说。

混元Hunyuan-MT-7B翻译：特朗普的共和党盟友及政府官员（包括副总统JD·万斯）纷纷在社交媒体上批评那些严厉质问卫生部长的议员。万斯在X平台上表示：“你们满口胡言，大家都清楚这一点。”

混元Hunyuan-MT-7B在翻译细节上的处理让我很惊喜，主谓很清晰，读起来完全不拗口。而且，这波在信达雅上，也是略胜一筹了。

2、俚语不别扭

能翻译明白俚语，绝对算得上进阶技能。

Hunyuan-MT-7B

长难句

原文（摘自路透社）：

Trump's Republican allies and members of the administration, including Vice President JD Vance, took to social media to criticize **lawmakers who had grilled the health secretary.** "You're full of shit and everyone knows it," Vance said on X.

页面翻译：

特朗普的共和党盟友和政府成员，包括副总统 JD Vance，在社交媒体上批评了**抨击这位卫生部长的立法者**。“**你满身都是狗屎**，每个人都知道，”万斯在 X 上说。

混元Hunyuan-MT-7B翻译：

特朗普的共和党盟友及政府官员（包括副总统JD·万斯）纷纷在社交媒体上批评**那些严厉质问卫生部长**的**议员**。万斯在X平台上表示：“**你们满口胡言**，大家都清楚这一点。”

俚语

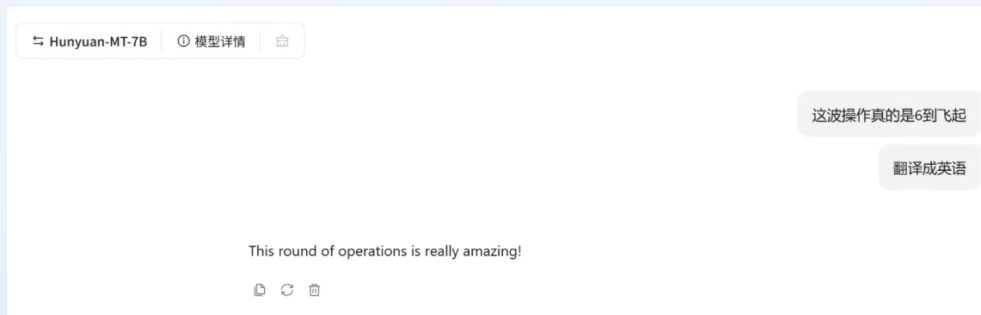
原文（摘自商业内幕）：**No matter how you slice it**, the August jobs report showed a weakening labor market.

页面翻译：

不管你怎么切片，8月份的就业报告都显示劳动力市场疲软。

混元Hunyuan-MT-7B翻译：

无论从哪个角度来看，8月份的就业报告都显示劳动力市场正在疲软。



Hunyuan-MT-7B

Hunyuan-MT-7B 模型详情

独独子立 沆瀣一气 踽踽独行 醍醐灌顶

翻译成英语

Alone and desolate; united as one; walking alone in solitude; as if being enlightened by a sudden burst of wisdom

公众号 · Simonlin的精神世界

Hunyuan-MT-7B 模型详情

三个臭皮匠，顶个诸葛亮

翻译成英语

Three cobblers with their wits combined can match Zhuge Liang.

公众号 · Simonlin的精神世界

Hunyuan-MT-7B 模型详情

大语言模型、自回归、耦合性、向量化、自注意力机制

翻译成英语

Large language models, autoregression, coupling, vectorization, self-attention mechanisms

公众号 · Simonlin的精神世界

Don't have a cow

翻译成简体中文

别生气；别大惊小怪

公众号 · 赛博小熊猫Loki

break a leg

翻译成简体中文

祝你好运；祝你能成功

公众号 · 赛博小熊猫Loki

都知道小红薯在国外风靡了，数据来告诉你有多逆天！

翻译成英语

Everyone knows that REDnote has become incredibly popular abroad. The data will show you just how amazing its success is!

没有一个中国人可以逃脱拼多多砍一刀。

翻译成英语

No Chinese person can escape the price cuts from Pinduoduo.

公众号 · 好好聊AI

have a monkey on my back

翻译成简体中文

（因担心或焦虑而）坐立不安

公众号 · 赛博小熊猫Loki

Hunyuan-MT-Chimera-7B

Hunyuan-MT-Chimera-7B

Analyze the following multiple Chinese translations of the Japanese segment surrounded in triple backslashes and generate a single refined Chinese translation. Only output the refined translation, do not explain.

The Japanese segment:

そこへ、監督の別府さんがたをあらわした。選手たちは、別府さんのまわりに集まって、めいめい、ぼうしをぬいで、あいさつをした。

The multiple Chinese translations:

1. 这时，教练别府先生出现了。选手们聚集在别府先生周围，各自脱下帽子，向他问好。
2. 这时，别府教练现身了。队员们纷纷聚集到他身边，各自脱帽致意。
3. 就在这时，别府教练出现了。选手们聚集在别府教练的身边，各自脱下帽子，打了个招呼。
4. 这时，教练别府先生出现了。选手们聚集到别府先生身边，各自脱下帽子，向他问好。
5. 就在这时，监督别府出现了。选手们纷纷聚集到别府身边，各自摘下帽子，向他问好。
6. 就在这时，教练别府先生出现了。队员们聚集在别府先生周围，纷纷摘下帽子，向他打招呼。

这时，教练别府出现了。队员们纷纷聚集到别府身边，各自摘下帽子向他问好。

Hunyuan-MT-Chimera-7B

Analyze the following multiple Chinese translations of the English segment surrounded in triple backslashes and generate a single refined Chinese translation. Only output the refined translation, do not explain.

The English segment:

I recently returned to d2 after several year, now I'm wondering:Where do you guys sell/buy your stuff? Do you just make a game "O xxx N yyy" and hope for the best? Or are there a website that's more efficient?

The multiple Chinese translations:

1. 我最近在几年后重新回到了d2，现在我知道：你们在哪里买卖东西？你们是直接创建一个"O xxx N yyy"的游戏然后碰运气吗？还是有更高效的网站？
2. 我最近几年后又回到了d2，现在我在想：你们一般在哪儿买卖装备？你们只是开个游戏房间"O xxx N yyy"，然后碰运气吗？还是有更高效的网站？
3. 时隔多年重回暗黑2，现在我知道：你们都在哪买卖装备？是直接建个"O物品换N符文"的房间碰运气吗？还是有更高效的交易网站？
4. 我最近在离开几年后重新回到d2，现在我很疑惑：你们都是在哪里买卖东西的？你们只是做一款游戏叫"O xxx N yyy"，然后听天由命吗？还是有更高效的网站？
5. 我最近时隔几年又回来玩d2了，想问一下：各位都是在哪里交易物品的？是自己建个名叫"出xxx换yyy"的房间然后随缘等，还是有更高效的网站？
6. 我最近在几年后回到了d2，现在我在想：你们这些人在哪里卖/买你们的东西？你们只是做一个游戏"O xxx N yyy"然后希望是最好的吗？还是有一个更有效率的网站？

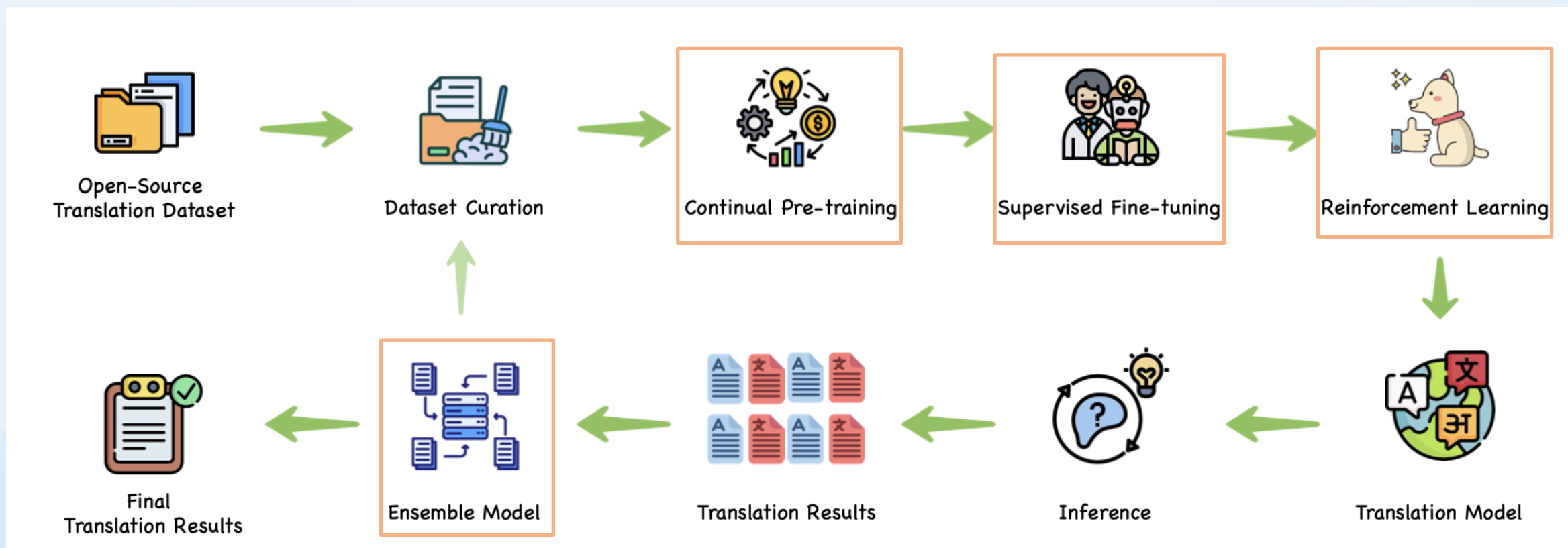
时隔多年后，我最近重新回归到了《暗黑破坏神2》(D2)。现在我知道：大家通常是在哪里买卖装备的呢？是直接创建一个类似"O xxx N yyy"这样的交易房间然后碰运气吗？还是存在更高效、更方便的交易网站呢？

字数: 104 | 输入tokens: 395 | 输出tokens: 457 | 总耗时: 1.19s

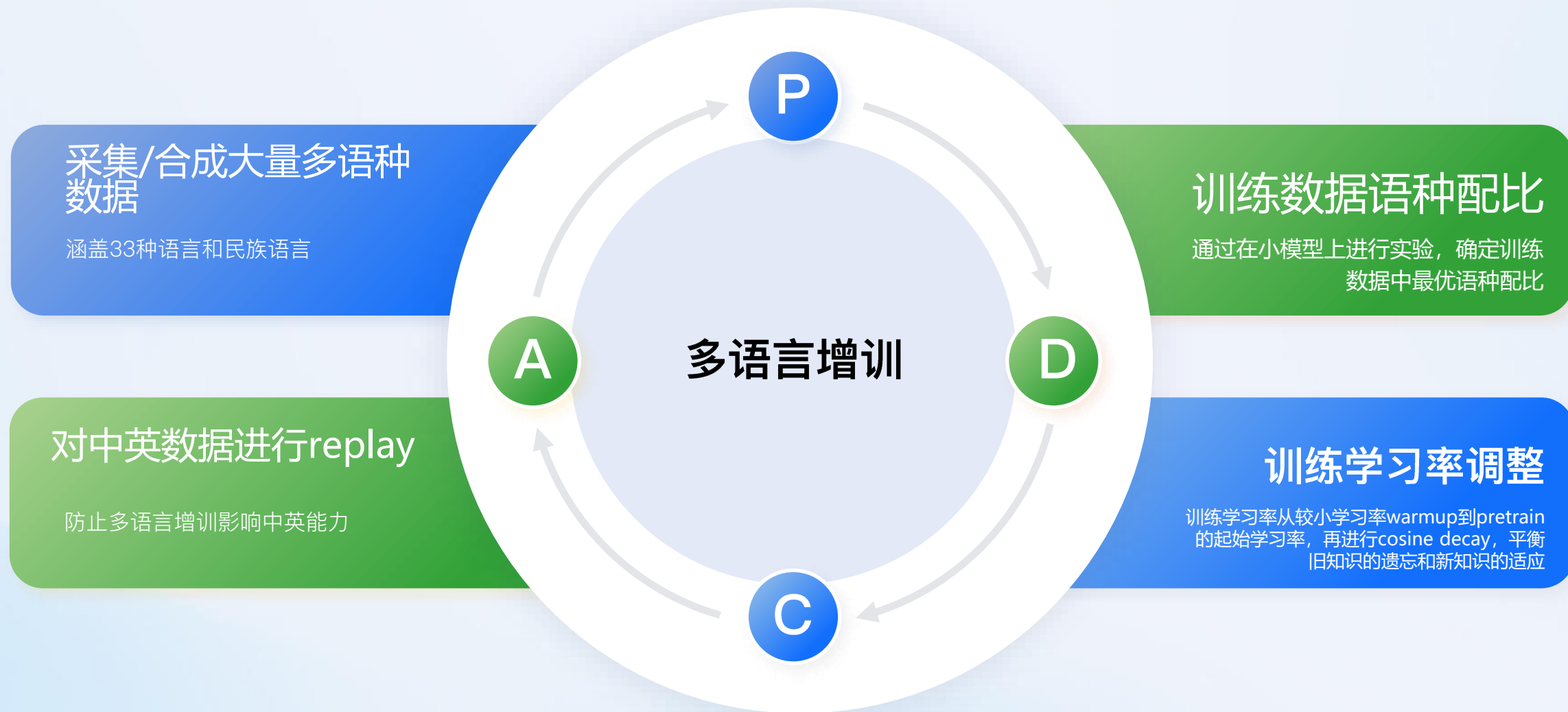
03

HunYuan-MT模型训练方案详解

Shy训练框架概览 (WMT方案)



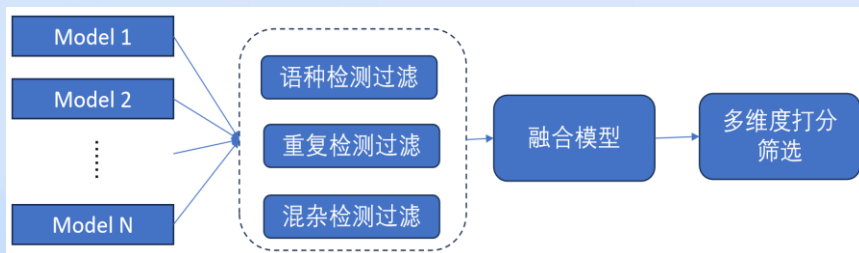
● 如何进行多语言增训 (CPT)



如何获取高质量平行语料 (SFT)

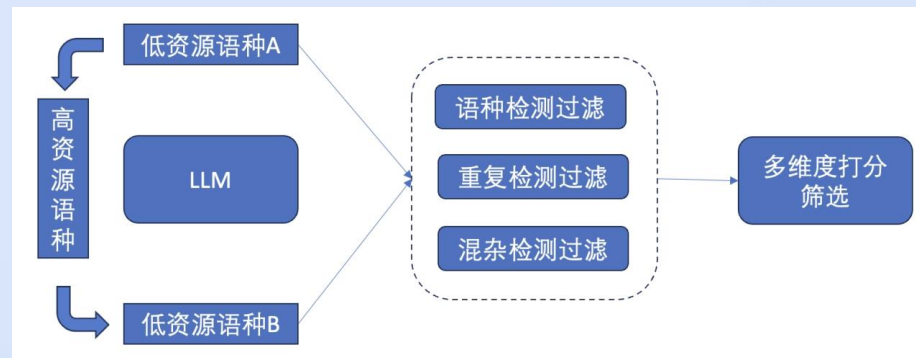
高资源语种

先采用少量网络收集和内部人工标注的高质量数据训练一个融合模型，然后使用该融合模型，将多个效果领先的大模型翻译结果进行融合，最终得到质量更高的翻译结果

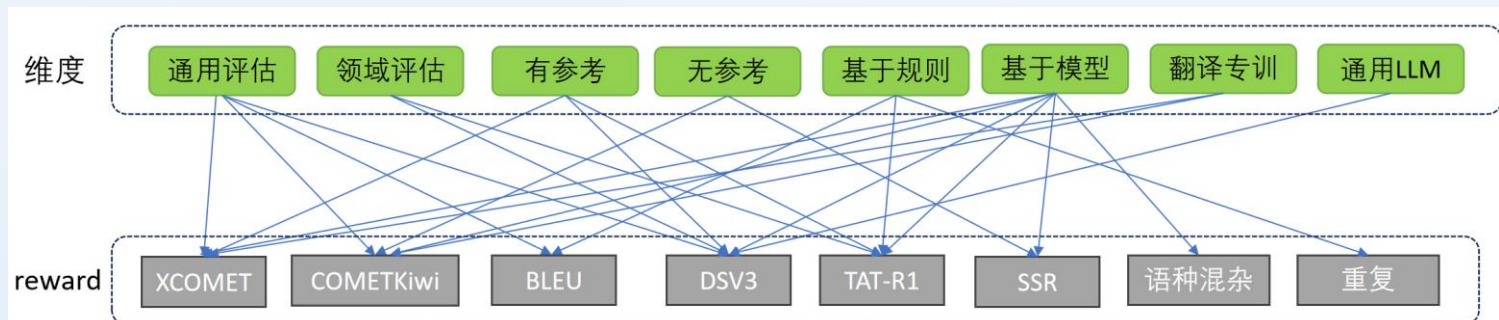


低资源语种

对于低资源语种，采用高资源语种（如英语）进行桥接得到低资源语种之间的平行语料，再通过多维度质量打分进行筛选



如何设计多语言RM (RL)





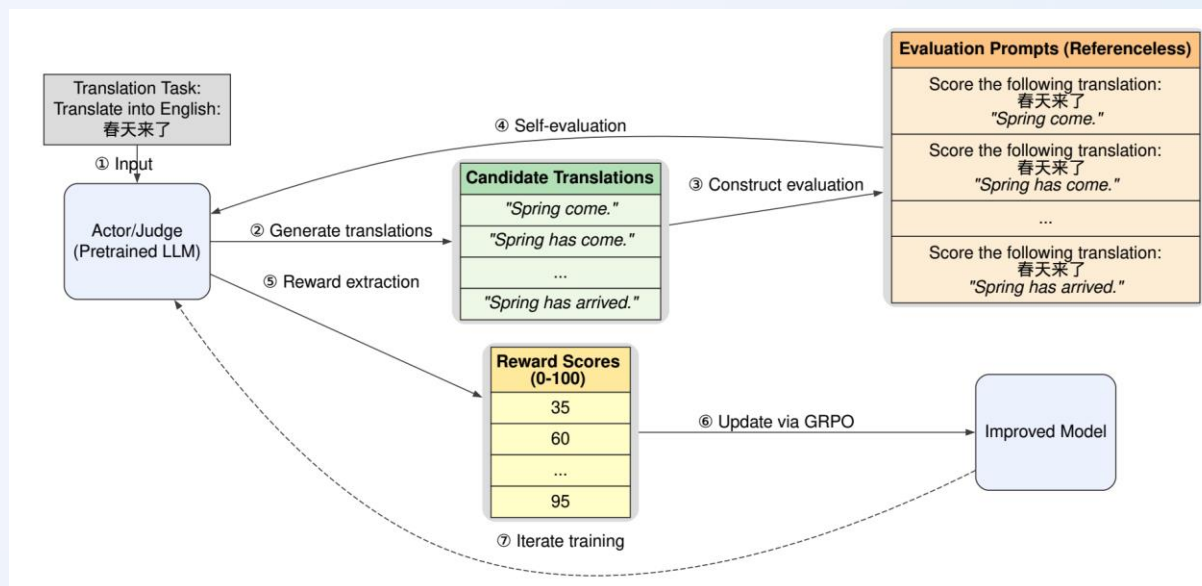
如何提升低资源语种翻译效果 (RL)

SSR (Simple Self-Rewarding) 强化学习训练方案

利用回译方法，只依赖单语数据即可训练。

让模型把语种A文本翻译成语种B，再翻译回语种A，通过计算两个语种A文本的相似性来得到reward信号。

使用特定prompt对翻译质量进行打分，最终使用该打分进行GRPO强化学习训练。这样既能做到不依赖平行语料，又能做到翻译能力和翻译打分能力之间的相互促进。



Models	ZH→EN				EN→ZH			
	WMT23	WMT24	Flores200	Avg.	WMT23	WMT24	Flores200	Avg.
	KIWI	XCM	KIWI	XCM	KIWI	XCM	KIWI	XCM
Closed-Source LLMs								
Claude-3.5-Sonnet	81.61	93.06	81.06	90.54	89.41	97.68	88.89	
GPT-4o	80.92	92.15	79.90	89.06	88.94	96.50	87.91	
Gemini-1.5-Pro	80.71	92.44	79.02	88.90	88.15	97.32	87.76	
Open-Source LLMs								
Qwen3-32B 🏆	79.74	90.79	79.20	88.47	87.68	95.75	86.94	
Qwen3-32B	80.28	91.95	79.95	89.53	88.88	97.18	87.96	
Qwen3-8B 🏆	78.30	89.03	77.99	86.94	85.82	93.89	85.33	
Qwen3-8B	79.87	91.42	79.58	89.02	88.61	96.55	87.51	
Qwen2.5-72B-Instruct	80.62	92.14	80.46	90.06	88.90	97.28	88.24	
Qwen2.5-32B-Instruct	77.73	89.28	78.77	88.69	87.13	95.50	86.18	
Qwen2.5-7B-Instruct	77.56	89.40	76.71	87.12	86.28	94.06	85.19	
QwQ-32B 🏆	74.61	85.12	75.08	84.34	80.88	89.21	81.54	
Gemma2-27B-it	80.32	91.96	79.42	89.14	88.64	96.72	87.70	
Gemma2-9B-it	79.86	91.21	79.25	88.41	88.32	96.25	87.22	
MT-Specific LLMs								
TowerInstruct-7B-v0.2	77.78	89.13	76.96	85.98	86.95	94.88	85.28	
TowerInstruct-13B-v0.1	78.53	89.90	77.57	87.12	87.30	95.80	86.04	
DeepTrans-7B 🏆	/	/	/	/	/	/	/	
Gemma2-28-9B-v0.1	79.40	90.63	78.71	88.60	87.85	96.33	86.92	
Ours								
Qwen2.5-7B	62.62	75.69	69.04	77.33	73.62	85.54	73.97	
SSR-Zero-7B 🏆	79.29	92.04	79.04	89.19	87.97	96.70	87.37	
SSR-X-Zero-7B 🏆	80.62	91.92	80.56	89.42	88.84	96.62	88.00	

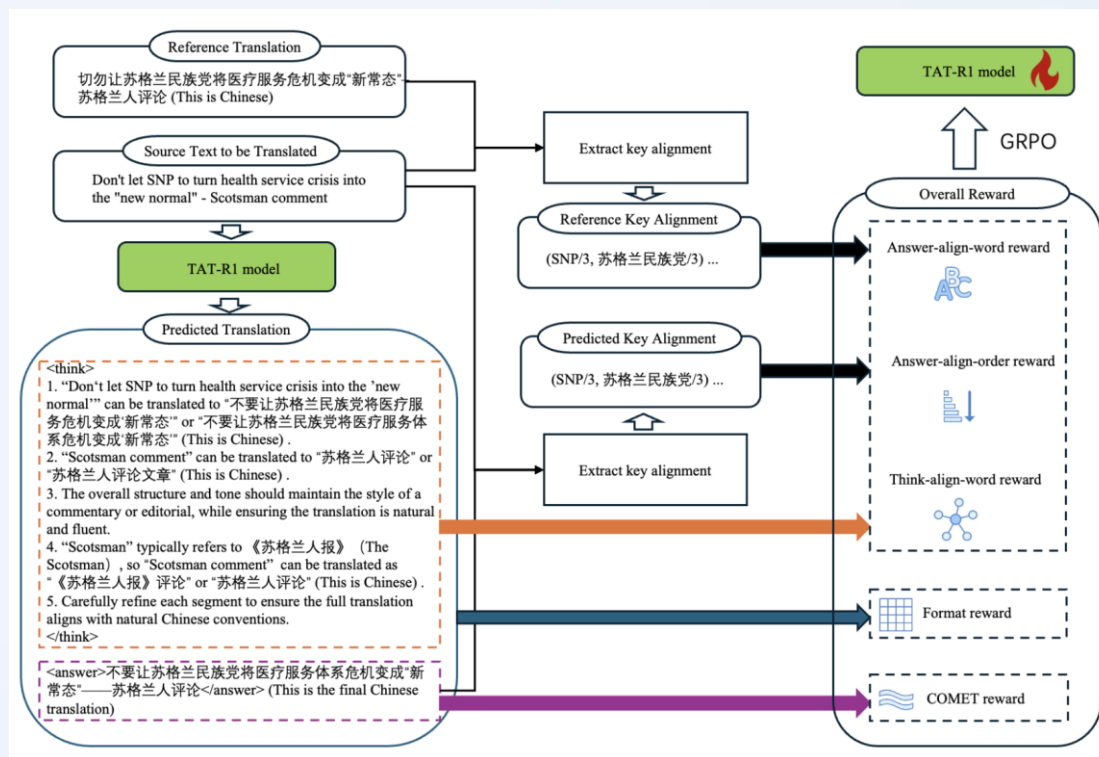
低资源语种效果达到同尺寸SOTA

<https://arxiv.org/pdf/2505.16637>

如何提升术语翻译效果 (RL)

TAT-R1训练方法

在强化学习训练阶段，提出TAT-R1训练方法，在RM中引入术语词对齐、术语出现顺序、思维链是否包含关键术语等维度



Models	EN->DE				
	BLEU	COMETKiwi	XCOMET	TA	Avg.
Baseline	25.87	67.05	88.65	53.29	58.72
NER + prompt	25.93	67.11	88.74	53.35	58.78
RL- R_{comet}	24.52	70.26	90.17	54.42	59.84
RL- $R_{comet}+R_{BLEU}$	27.37	66.49	88.77	54.91	59.39
RL- $R_{comet}+R_{aaw}$	26.21	71.33	90.56	55.57	60.92
RL- $R_{comet}+R_{aaw}+R_{aao}$	26.34	72.04	90.99	55.73	61.28
RL- R_{all} (TAT-R1)	27.10	73.82	91.22	56.42	62.14

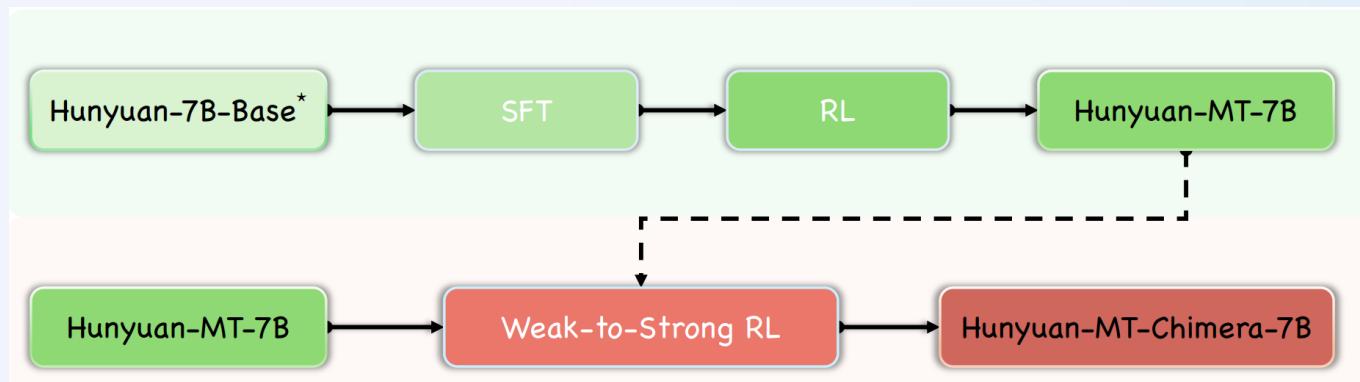
通用翻译能力不下降的情况下，模型术语翻译效果大幅度提升

<https://arxiv.org/pdf/2505.21172>

Hunyuan-MT-Chimera (RL)

打造有参考翻译能力 (Translation with Reference)

在实际应用场景中，存在大量有参考翻译能力的需求，但是模型侧对于有参考翻译能力属于空白。



Example #1	I recently returned to d2 after several year, now I'm wondering: Where do you guys sell/buy your stuff? Do you just make a game "O xxx N yyy" and hope for the best? Or are there a website that's more efficient?
Testing Points	"d2": Diablo II, "make a game "O xxx N yyy"": it means creating a game room where the room name uses "O" (short for "Offer," standing for the item xxx you want to sell) and "N" (short for "Need," standing for the item yyy you want in exchange) to inform other players of your trading needs.
Model #1	我最近几年后又回到了d2，现在我在想：你们一般在哪里买卖装备？你们只是开个游戏房间“O xxx N yyy”，然后碰运气吗？还是有更高效的网站？
Model #2	我最近在几年后重新回到了d2，现在我想知道：你们都在哪里买卖东西？你们只是创建一个“O xxx N yyy”的游戏然后碰运气吗？还是有更高效的网站？
Model #3	我最近时隔几年又回来玩d2了，想问一下：各位都是在哪里交易物品的？是自己建个名叫“出xxx换yyy”的房间然后随缘等，还是有更高效的网站？
Model #4	时隔多年重回暗黑2，现在我想知道：你们都在哪里买卖装备？是直接建个“O物品换N符文”的房间碰运气吗？还是有更高效的交易网站？
Model #5	我最近在离开几年后重新回到d2，现在我很疑惑：你们都是在哪里买卖东西的？你们只是做一款游戏叫“O xxx N yyy”，然后听天由命吗？还是有更高效的网站？
Model #6	我最近在几年后回到了d2，现在我在想：你们这些人在哪里卖/买你们的東西？你们只是做一个游戏“O xxx N yyy”然后希望是最好的吗？还是有一个更有效率的网站？
Hunyuan-MT-Chimera-7B	时隔多年，我又重新开始玩《暗黑破坏神2》（Diablo II），现在我想知道：大家都是在哪儿进行物品交易的？是通过自己创建名为“O xxx N yyy”的游戏房间来交易，然后听天由命吗？还是有更高效的交易网站？

Example #4

Model #1

一辆1948年的福特水星汽车穿过佐治亚州门罗小镇的一群围观者，朝着小小的摩尔滩桥隆隆奔行。

A 1948 Ford Mercury **drove through** a crowd of onlookers in the town of Monroe, Georgia, and then made its way **across** the small Moorestown Bridge.

Model #2

A 1948 Ford Mercury **drove through** a crowd of onlookers in the town of Monroe, Georgia, and then made its way **across** the small Moore's Bay Bridge.

Model #3

A 1948 Ford Mercury **convertible drove through** a crowd of onlookers in Monroe, Georgia, and then made its way **across** the small Moore's Bend Bridge.

Model #4

A 1948 Ford Mercury **drove through** a crowd of onlookers in the town of Monroe, Georgia, and then made its way **across** the small Moorestown Bridge.

Model #5

A 1948 Ford Mercury **driven by a man named George White** sped through a crowd in Monroe, Georgia, and then over the small Moore's Bay Bridge.

Model #6

A 1948 Ford Mercury **driven by a man named Steve Myers** sped through a crowd in Monroe, Georgia, and over the small bridge that crosses Moore Creek.

Hunyuan-MT-Chimera-7B

A 1948 Ford Mercury **sped through** a crowd of onlookers in the small town of Monroe, Georgia, **on its way towards** the tiny Moorestown Bridge.

04

Hunyun-MT应用场景能力建设方案

应用场景能力建设

带格式翻译

典型场景：
企微邮件翻译，QQ浏览器网页翻译，腾讯视频字幕翻译。

实时翻译

典型场景：
腾讯会议中翻译，腾讯财报会同声传译，QQ浏览器视频/直播翻译



术语、例句库翻译

典型场景：
定制化翻译需求的场景、领域翻译专业度要求高

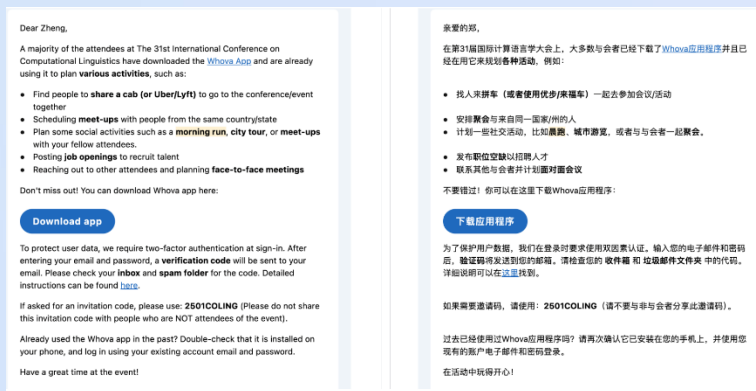
词典翻译&解释性翻译

典型场景：
划词翻译、教育场景倡议能力



面相应用场景能力建设-1

带格式翻译



解决方案：

- 设计一协议承载原始格式信息，可以多个相邻文本一起翻译，保证翻译语义套连贯性。
- 借助LLM和词对齐等工具构造训练数据，并人工check确保数据质量
- SFT进行冷启+GRPO强化提升效果，reward既关注翻译质量也关注格式信息保留效果

实时翻译



解决方案：

- 模型小型化，采用更小激活参数模型
- 补充口语化数据，通过强化学习不断提升效果
- 重点增强多轮对话场景的模型上下文翻译一致性



面相应用场景能力建设-2

术语、例句库翻译

Spring具有以下以下含义，请根据上下文选择合适的翻译：

1.春天

Spring is the season of new beginnings. (春天是新的开始的季节。)

2.弹簧

The mattress has a lot of springs. (床垫有很多弹簧。)

请把下面的文本翻译成中文

The spring in the pen is broken.

笔里的弹簧坏了。

字数: 8 | 输入to

Spring具有以下以下含义，请根据上下文选择合适的翻译：

1.春天

Spring is the season of new beginnings. (春天是新的开始的季节。)

2.弹簧

The mattress has a lot of springs. (床垫有很多弹簧。)

请把下面的文本翻译成中文

I love the flowers in spring.

我喜欢春天的花。

解决方案：

- 建设翻译系统，提供可编辑、实时生效的术语、例句库检索模块
- 提升模型的术语指令遵循能力，适应多种语境

词典翻译&解释性翻译

cryptocurrency

英 /'krɪptəʊkʌrənsi/ 美 /'krɪptəʊkɜːrənsi/

n. 加密货币（一种基于密码学技术的数字资产，不依赖中央银行或政府发行，如比特币、以太坊等）

compliance

英 /kəm'plaɪəns/ 美 /kəm'plaɪəns/

n. 服从；遵守；合规；符合（标准或要求）

翻译：

英雄所见略同。

解释：

“Great minds think alike” 是一句常用英语谚语，字面意为“伟大的头脑想法相似”，用于表达两人或多人在某个问题上不谋而合，有着相同的看法或思路，是对彼此共识的一种友好且略带夸赞的表达。

解决方案：

- 词典知识注入：收集词典数据，如牛津词典、21世纪大英汉词典等
- 例句解释知识注入：解释性翻译数据收集

05

经验总结

关键经验总结

高质量数据获取和构造是重中之重

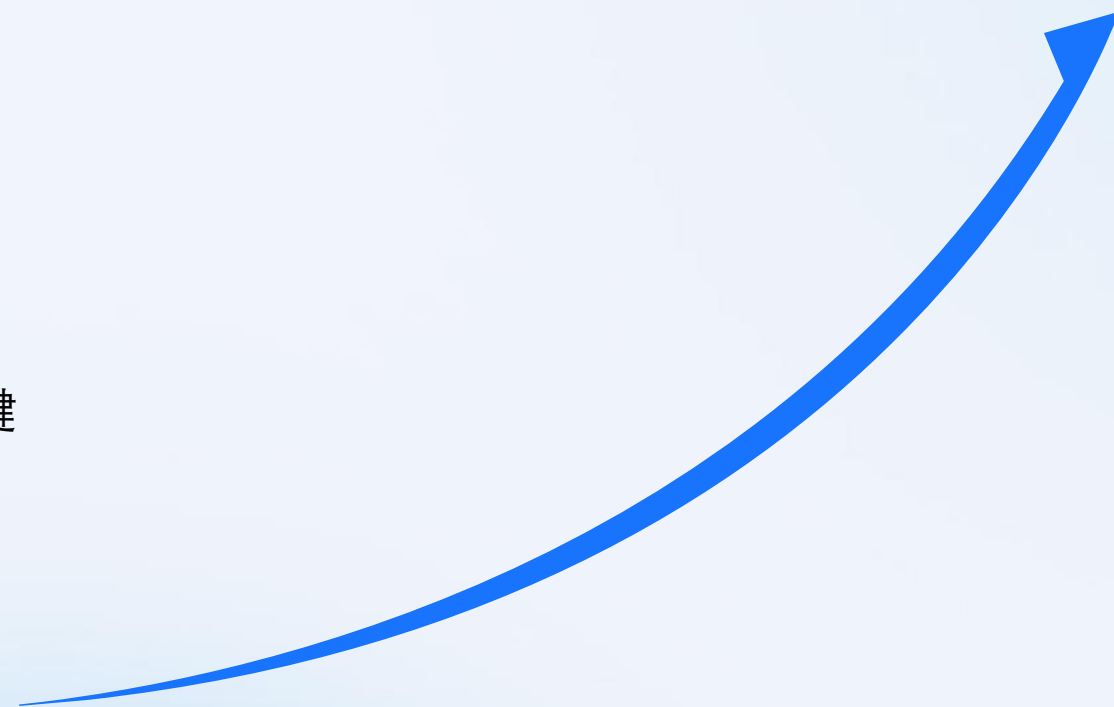
- 离线构造和获取数据花费精力最多，时间最长
- 构造数据环节可以做很多模型优化的工作，辅助提升数据质量

Reward Model如何定义清晰非常有挑战

- 信达雅是否足够
- 多样化指令遵循如何满足

强化学习阶段要稳定、长时间训练

- 强化学习算法的优化不可小觑，是拉开效果差距的关键
- 训练效率很关键



06

未来展望

Hunyuan-MT未来展望

DENSE模型架构升级



1

MOE模型架构

记住更多词汇、支持更长上下文

文本翻译



2

多模态翻译

语音端到端翻译、图片/视频翻译

翻译模型



3

翻译AGENT

用memory解决长文翻译一致性问题

📍 北京

QCon

全球软件开发大会

会议时间：4月10-12日

- 大模型赋能 AIOps
- 越挫越勇的大前端
- 云上业务架构演进
- 多模态大模型及应用
- 海外 AI 应用创新实践

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：6月27-28日

- 端侧智能
- AI Agent
- 多模态大模型
- 金融+大模型

📍 上海

QCon

全球软件开发大会

会议时间：10月23-25日

- AI Agent
- 大模型训练推理
- 端侧 AI
- 搜推深度融合
- 数智金融

4月

5月

6月

8月

10月

12月

📍 上海

AiCon

全球人工智能开发与应用大会

会议时间：5月23-24日

- 通用大模型
- AI Agent
- 垂直领域应用
- 模型可解释性

📍 深圳

AiCon

全球人工智能开发与应用大会

会议时间：8月22-23日

- 模型效率与部署
- 多模态大模型
- 大模型安全
- 智能硬件

📍 北京

AiCon

全球人工智能开发与应用大会

会议时间：12月19-20日

- 通用大模型
- 智能硬件
- LMOPs
- 具身智能

极客邦科技 2025 年会议规划

促进软件开发及相关领域知识与创新的传播



参会咨询



查看会议

QCon
全球软件开发大会

THANKS

大模型正在重新定义软件

Large Language Model Is Redefining The Software

微信

