

Large Language Models and Machine Translation Research

Chengqing Zong

Institute of Automation, Chinese Academy of Sciences

cqzong@nlpr.ia.ac.cn

Outline

- 
- 1. Introduction**
 - 2. Performance analysis of translation by LLMs**
 - 3. What we should do?**
 - 4. Some recent work on MT**
 - 5. Conclusion**

1. Introduction

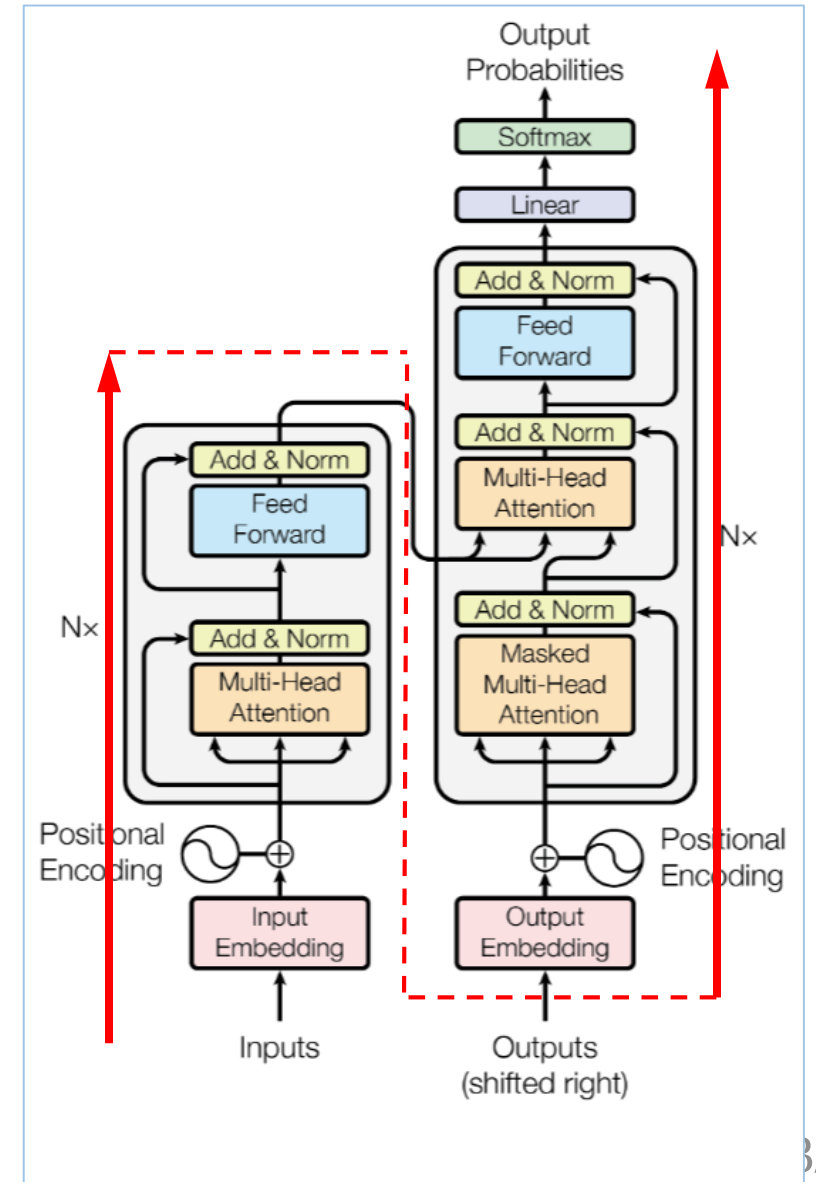
◆ Attention and Transformer

A. Vaswani et al., [Attention Is All You Need](#).
Proc. of NIPS'2017, pp 6000–6010

◆ Pre-trained language models

- ELMo [Peters et al., 2018] (AllenAI&WU)
- BERT [Devlin et al., 2019] (Google)
- GPT [Radford et al., 2018] (OpenAI)
- **G**enerative **P**re-**T**rained Transformer

According to the definition by Wikipedia, language models with a parameter scale exceeding one billion are referred to as **large language models (LLMs)**.



1. Introduction

- GPT-3.5, Nov. 29, 2022
- GPT-4.0, March 14, 2023
- Sora, Feb. 15, 2024
- GPT-4o (Omni), May 13, 2024
- o1-preview, o1-mini, Sept. 12, 2024
- o3, o3 mini, Dec. 20, 2024
- Deepseek-R1, Jan. 20, 2025



Prompt: A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.

1. Introduction

Given a sentence in a language $\Rightarrow$ A sentence in another language (MT)

Given a question $\Rightarrow$ The answer to the question (Q&A)

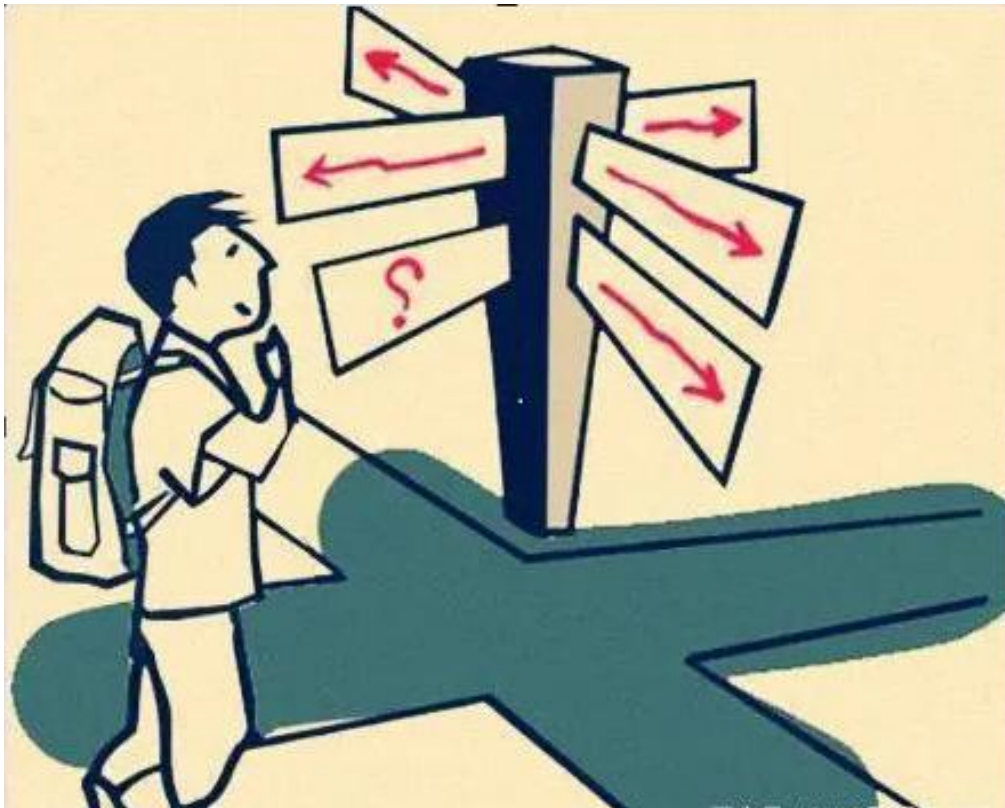
- Machine translation
- Human-dialogue system
- Q&A
- Sentiment/ emotion analysis
- Text classification
- Automatic summarization
- Automatic writing / language generation
- Speech recognition
- Image caption
-

Artificial General Intelligence
— AGI

1. Introduction

● Questions:

- How do the LLMs perform in practical tasks, such as machine translation (MT)?
- How will NLP technologies, including MT, develop in the future?



To be or not to be, this is a question.

Outline

1. Introduction
- ➔ 2. **Performance analysis of translation by LLMs**
3. **What we should do?**
4. **Some recent work on MT**
5. **Conclusion**



2. Performance analysis of translation by LLMs

◆ Objective evaluation of MT performance (May 2023)

- (1) Rich resource languages: Chinese, German
- (2) Medium resource languages: Estonian (爱沙尼亚语), Lithuanian (立陶宛语)
- (3) Low resource languages: Sinhalese (僧伽罗语), Nepali (尼泊尔语)

Test data:
Flores-200

Language pairs	COMET (%)			BLEU		
	ChatGPT	Google	Ratio	ChatGPT	Google	Ratio (%)
Chinese → English	88.7	89.4	99.3	29.6	38.6	76.7
German → English	90.7	91.0	99.7	43.0	47.3	90.9
Estonian → English	90.1	91.7	98.3	36.9	45.5	81.1
Lithuanian → English	86.1	89.2	96.6	32.6	41.3	78.9
Nepali → English	87.4	93.2	93.8	24.5	51.3	47.7
Sinhalese → English	61.3	90.5	67.7	3.5	45.5	7.7



2. Performance analysis of translation by LLMs

➤ Ability to add world knowledge

Models	Examples
Source	On Dec. 17, according to disclosure documents, Paul Pelosi bought 10 Alphabet call options with a strike price of \$2,000 with an expiration date set for mid-September 2022.
ChatGPT	根据披露文件，保罗·佩洛西于12月17日购买了10份 谷歌母公司 的认购期权，行使价格为2000美元，到期日定于2022年9月中旬。
Google Translate	根据披露文件，12月17日，保罗·佩洛西 (Paul Pelosi) 购买了 10 份 Alphabet 看涨期权，行使价为 2,000 美元，到期日定为 2022 年 9 月中旬。
Source	Visa was the sole Dow component to decline, with the other 29 stocks in the benchmark index finishing the day in the green .
ChatGPT	Visa是唯一一只道琼斯指数下跌的成分股，其他29只股票在基准指数中 收于涨势 。
Google Translate	Visa 是唯一一下跌的道琼斯指数成份股，该基准指数中的其他 29 只股票当天 以绿色收盘 。

J. Wu et al., Capability Analysis and Future Prospects of ChatGPT, China Science Foundation, Vol. 37, No. 5, 2023, pages 735-742 (武俊宏等, ChatGPT 能力分析与未来展望, 中国科学基金, 2023年05期)



2. Performance analysis of translation by LLMs

● Comparison of translations by GPT and MS Translator

Sequence Type	Translation Instance	Phenomenon
Source	Bis auf die E 95 02 wurden <u>alle Lokomotiven zerlegt</u> .	Non-Monotonicity (NM)
MS Translator	With the exception of E 95 02, all locomotives were dismantled .	
GPT	<u>All locomotives were dismantled</u> except for the E 95 02.	
Source	Oder ist sie ganz aus dem Sortiment genommen?	Fluency (F)
MS Translator	Or is it completely removed from the range?	
GPT	<u>Or has it been</u> completely removed from the range?	
Source	Sehen Sie bitte im Screenshot was der Kollege geschrieben hat	Punctuation Insertion (PI)
MS Translator	Please see in the screenshot what the colleague wrote	
GPT	Please see the screenshot for what the colleague wrote.	
Source	Die Email zur Stornierung wurde am 26.12. <u>#NUMBER#</u> versendet.	Dropped Content (USW)
MS Translator	The cancellation email was sent on 26.12. <u>#NUMBER#</u> .	
GPT	The cancellation email was sent on December 26th.	
Source	"We won't accept the <u>CAA</u> and that is for sure.	Inserted Content (UTW)
MS Translator	“我们不会接受CAA，这是肯定的。	
GPT	“我们不会接受 <u>《公民法》</u> ，这是肯定的。	

Amr Hendy et al., Good Are GPT Models at Machine Translation? A Comprehensive Evaluation,
<https://arxiv.org/abs/2302.09210>



2. Performance analysis of translation by LLMs

● October 2024

➤ Data setting

- WMT23: German (de), Chinese (zh), Russian (ru), Japanese (ja), Ukrainian (uk), Hebrew (he)
- WMT22: Icelandic (is)

➤ Translation: $EN \rightarrow X$, $X \rightarrow EN$

	Rich Resources	Medium Resources	Low Resources
Same language family as English	de, ru	uk	is
Different language families from English	zh	ja	he



2. Performance analysis of translation by LLMs

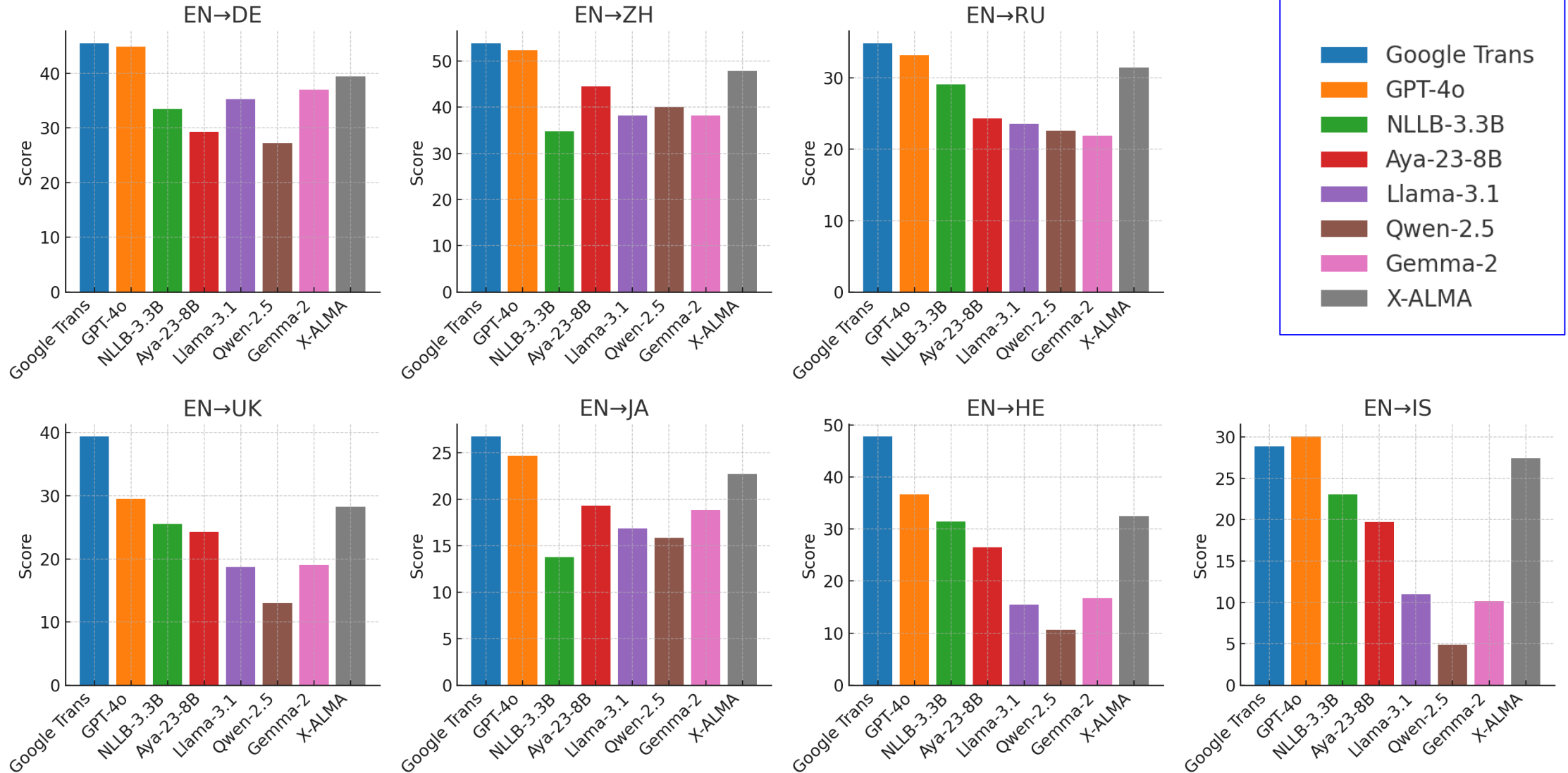
➤ Experimental setting

➤ Models/ Systems

- **API:** Google Translation, GPT-4o
- **Open Source Instruction Model:** NLLB-3.3B, Aya-23-8B, Llama-3.1-8B, Qwen-2.5-7B, Gemma-2-9B
- **Translation Model Based on LLM:** X-ALMA

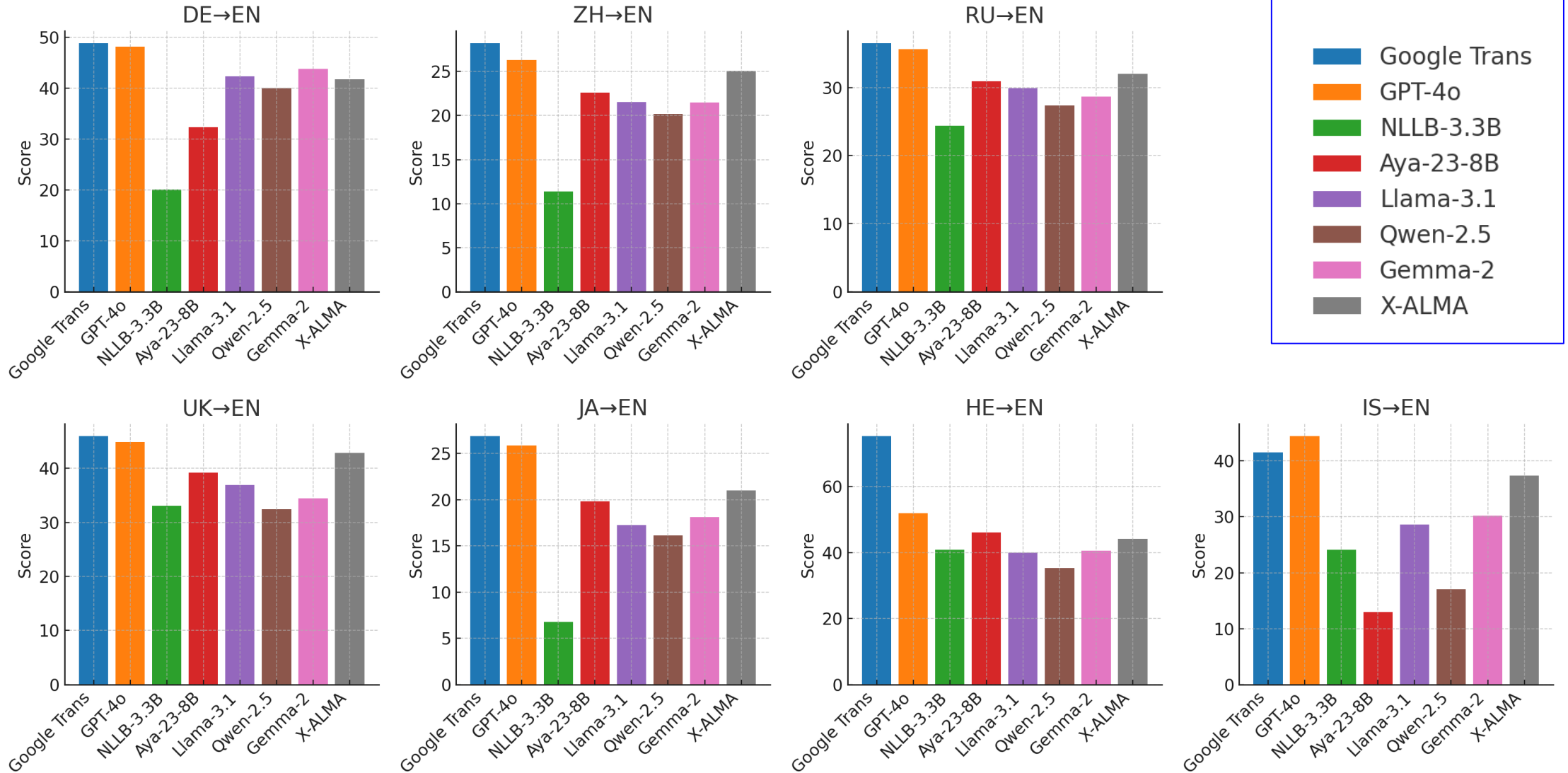
2. Performance analysis of translation by LLMs

B
L
E
U



2. Performance analysis of translation by LLMs

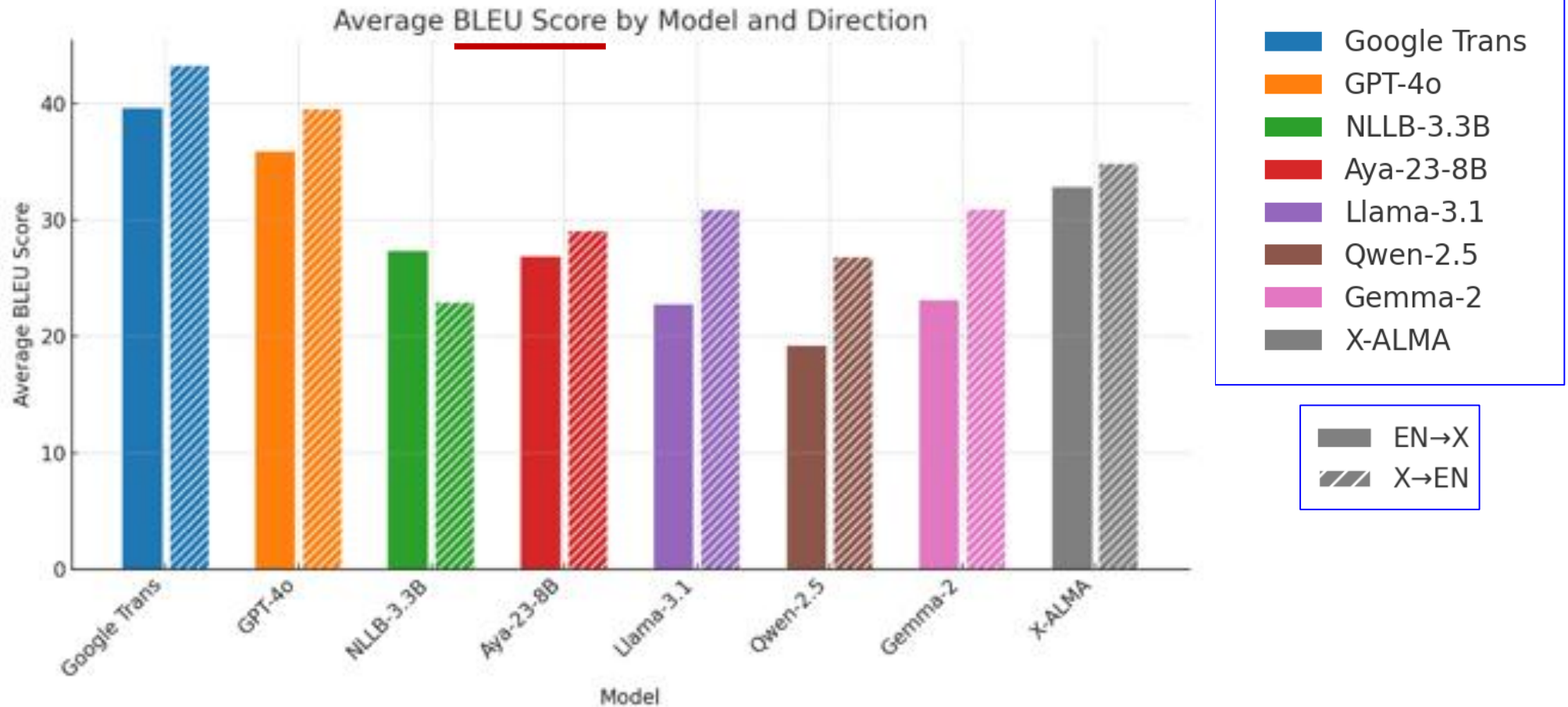
B
L
E
U





2. Performance analysis of translation by LLMs

● Average score





2. Performance analysis of translation by LLMs

● Summary

- The translation capability of GPT-4o is comparable to that of Google Translator;
- The open source models such as Llama-3.1 and Qwen-2.5 have already approached the performance of SOTA Multilingual Translation Model (NLLB) with Encoder-Decoder structure in terms of translation performance, but still have a certain gap compared to closed source models;
- The LLMs perform well in resource rich languages such as English, Chinese, and German; When dealing with low resource languages, LLMs can achieve relatively reasonable translation results with limited data through transfer learning and knowledge sharing, but still lag behind specially trained machine translation models based on Transformer architecture;



2. Performance analysis of translation by LLMs

- The LLMs perform well between languages of the same language family, such as English, German, and Russian in the Indo-European language family, but performs slightly worse between different language families, such as English and Chinese in the Sino-Tibetan language family;
- When dealing with translation tasks with non English as the target language ($EN \rightarrow X$), LLMs usually perform slightly weaker, which may be related to the bias towards English as the central language during training.



2. Performance analysis of translation by LLMs

◆ Subjective evaluation of MT performance and human translation quality

- For the undergraduate students, there is a major in translation in China universities. It is clearly not feasible to have a translation major without solid language skills. However, from the current situation of this major, the translation quality of undergraduate students is not as good as what machines can do.
- At the graduate stage, there is a Master of Translation and Interpretation (MTI) program, and the translation quality of graduates varies greatly. Many graduates' translation skills are not as good as those of MT systems.

In recent years, many universities have stopped admitting undergraduate students to English and some other foreign language majors.

Outline

1. Introduction
2. Performance analysis of translation by LLMs
- ➡ 3. **What we should do?**
4. **Some work on MT**
5. **Conclusion**



3. What we should do?

Q1: Can machine translator replace human interpreters? Yes / No ?

Example-1: 我们要传承红色基因，赓续红色血脉，肩负起中华民族的光荣使命。

MT-1(B): We must inherit the red genes, carry on the red bloodline, and shoulder the glorious mission of the Chinese nation.

MT-2(D): We must inherit the red genes, continue the red bloodline, and shoulder the glorious mission of the Chinese nation.

MT-3(K): We should inherit the red legacy, carry on the red bloodline, and fulfill the noble mission of the Chinese nation.

References: “红色基因” — traditions of revolution

“红色血脉” — revolutionary / communistic culture ?



3. What we should do?

Example-2: 黛玉自在枕上感念宝钗……又听见窗外竹梢焦叶之上，雨声淅沥，消寒透幕，不觉又滴下泪来。
(《红楼梦》第45回)

As she lay there alone, Dai-Yu's thoughts turned to Bao-chai ...Then she listened to the insistent rustle of the rain on the bamboos and plantains outside her window. The coldness penetrated the curtains of her bed. Almost without noticing it she had begun to cry.
— David Hawkes

Daiyu felt sorry for Baochai on the pillow..... **I heard** the sound of rain pouring on the scorched leaves of the bamboo shoots outside the window, and the cold was through the curtain, and **I didn't** realize that **I was** dripping tears again. (M, March 14, 2025)

Daiyu felt grateful for her **precious hairpin** on her pillow... Then she heard the sound of rain outside the window on the **charred leaves** of the bamboo shoots, dispelling the cold and penetrating the curtain. She couldn't help but shed tears again. (B, March 14, 2025)

3. What we should do?

- **Answer-1:**

For many important works such as literary classics that require high-quality translations, as well as important documents closely related to politics, diplomacy, culture, etc., MT system is not yet competent.

3. What we should do?

Q2: In practical application scenarios, is the performance of MT systems the same as the comparative analysis results in the lab.?

Someone has conducted statistics showing that **over 20%, 14%, and 10%** of the words in news texts in the mainstream media of Indonesian, Malay, and Filipino, respectively, have spelling errors,

For example, the word '**gagal**' in Indonesian, which means 'failure', is often replaced by 'ga2l'; In order to emphasize a certain tone or a mood, sometimes some letters are repeatedly inputted, e.g., 'cemangudh' is written as '**cemanguuuuuudh**'.

F is for fakes
Deepfakes are synthetic images, videos or audio in which someone is digitally altered so that they look, sound or act like someone else. Created by complex algorithms, deepfakes have raised concerns over their ability to spread misinformation.

G is for generative adversarial network
An AI technique that uses two competing computer systems to generate fake data that matches the real data as closely as possible. It can be used to create hyper-realistic deepfake photographs of people or animals who don't actually exist.

H is for human-in-the-loop
An AI system in which humans can intervene by training, tuning and testing the algorithm so that it produces more accurate results. It is a way of combining human and machine intelligence, helping to make up for the shortcomings of both.

I is for intelligence
A machine is said to have artificial intelligence if it's able to replicate some ability of humans or animals. That might be reasoning, learning from its mistakes, making decisions, communicating, solving problems, and/or moving around in the environment.

K is for Kasparov
In 1997, Gary Kasparov became the first person to lose a world chess championship to a computer.

L is for Lovelace
19th Century mathematician Ada Lovelace worked with Charles Babbage on his 'Analytical Engine' mechanical computer. It was never built, but she wrote a computer algorithm for it that many regard as the world's first, making her one of the first programmers.

J is for Julia
A modern programming language (launched in 2012) that is gaining in popularity among the AI and data science community because of its high speed, ease of use and flexibility. It is used in applications including climate modelling, robotics and cellular biology.

L is for Lovelace
19th Century mathematician Ada Lovelace worked with Charles Babbage on his 'Analytical Engine' mechanical computer. It was never built, but she wrote a computer algorithm for it that many regard as the world's first, making her one of the first programmers.

L is for Lovelace
19th Century mathematician Ada Lovelace worked with Charles Babbage on his 'Analytical Engine' mechanical computer. It was never built, but she wrote a computer algorithm for it that many regard as the world's first, making her one of the first programmers.

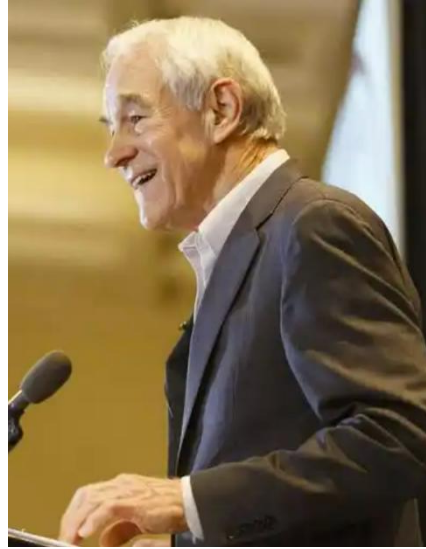
3. What we should do?

- **Answer-2:**

For texts with spelling errors, noise, and complex backgrounds in practical applications, MT systems still find it difficult to accurately and robustly deal with and translate them.

3. What we should do?

Q3: In practical applications, do MT systems only translate text?



3. What we should do?

- **Answer-3:**

At present, simultaneous interpretation technology is still difficult to have good results, especially in complex scenarios such as noisy venues, multiple speakers, and even debates, where the quality of simultaneous interpretation is far from meeting the required performance.

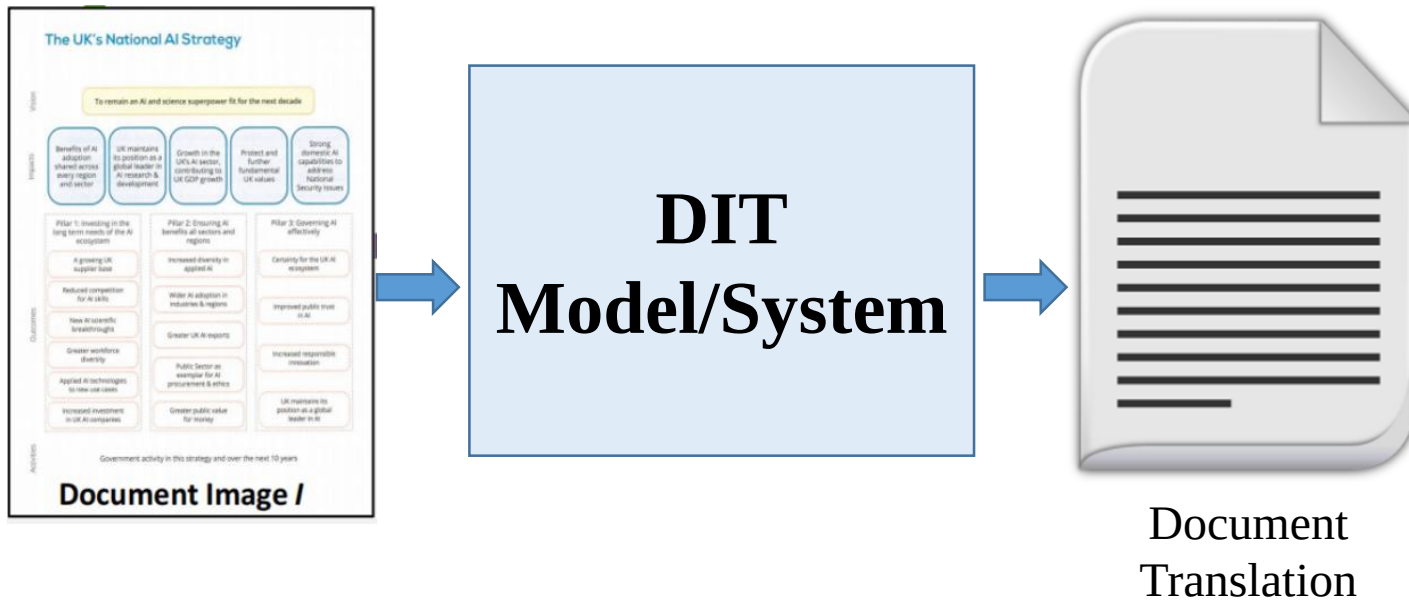
Outline

1. Introduction
2. Performance analysis of translation by LLMs
3. What we should do?
- ➡ 4. **Some work on MT**
5. **Conclusion**

4. Some recent work on MT

(1) Document Image Translation

The task of document image translation (DIT) is to perceive and recognize the text contained in the document image, and translate it from the source language to the specified target language.



$$\hat{Y} = \arg \max_Y P_{\theta} (Y | I, T, B)$$

$$= \arg \max_Y \prod_{j=1}^{|Y|} P_{\theta} (y_j | y_{<j}, I, T, B)$$

I - a given document image;

T - text recognized by OCR;

B - position of the text block recognized by OCR;

Y - the document of target language.

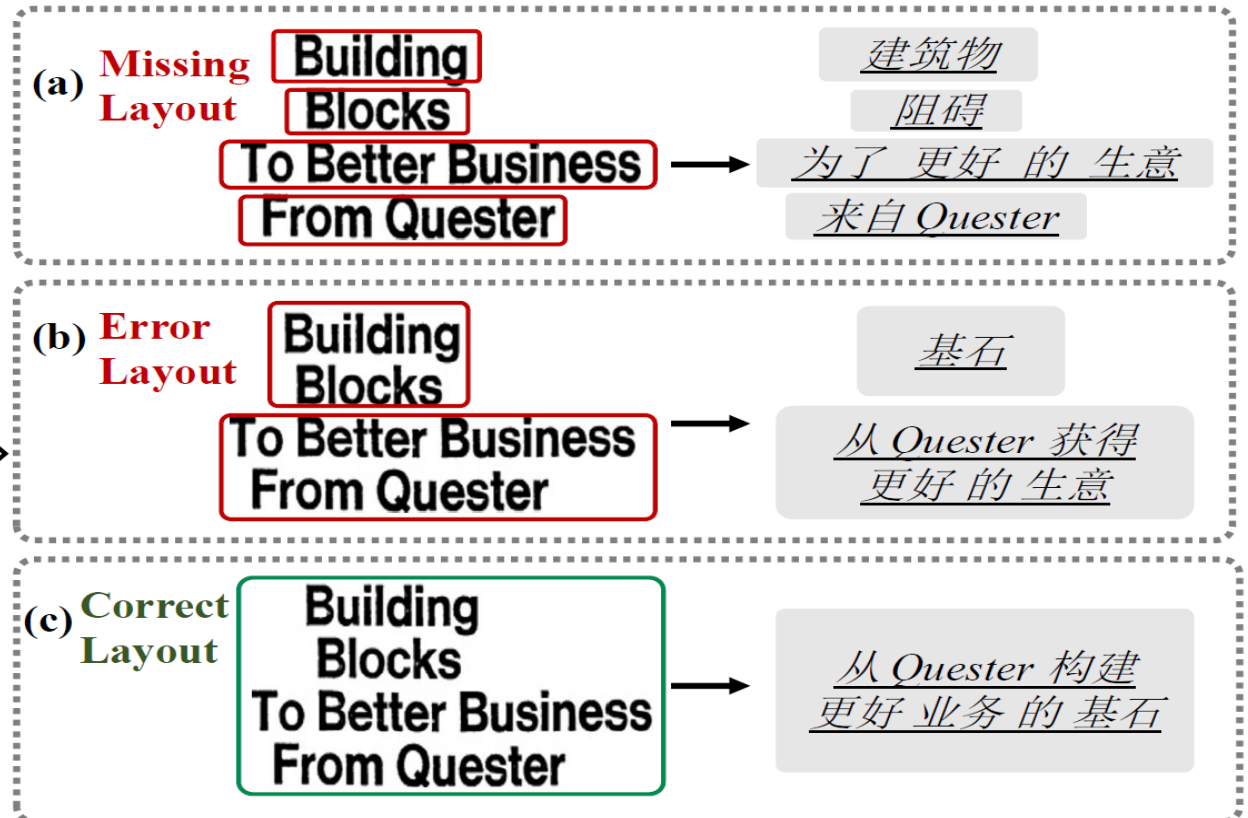
4. Some recent work on MT

● Layout information is very important for document image translation

- Neglecting layout information, Fig. (a), or misunderstanding, Fig. (b) will lead to wrong translation.



Building
Blocks
To Better Business
From Quester



Document Image

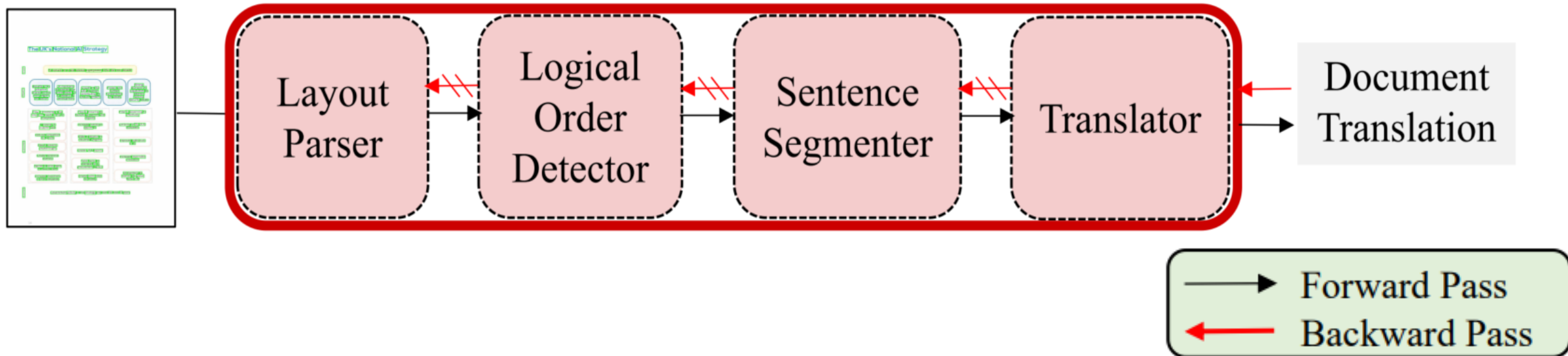
Raw OCR Results

OCR with Visual Layout

Translation Results

4. Some recent work on MT

- **Existing methods** — cascading translation using independent models for layout understanding and translation separately
 - **Advantages:** Simplify the modeling difficulty of various sub-tasks such as layout recognition and translation;
 - **Disadvantage:** The independent and non interactive modules may lead to insufficient utilization of the layout by the translator; The cascading paradigm carries the risk of error propagation accumulation.



4. Some recent work on MT

● Our method

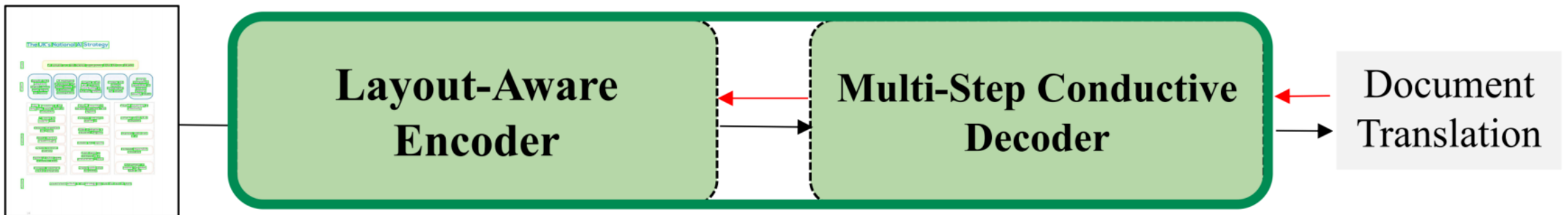
Two modules: a layout information aware encoder + a multi-step conduction decoder:

➤ **Layout information aware encoder (LIAE)**

- Extract and encode layout features, and integrate layout and text features;
- Perform deep contextual fusion on the fused layout text joint features.

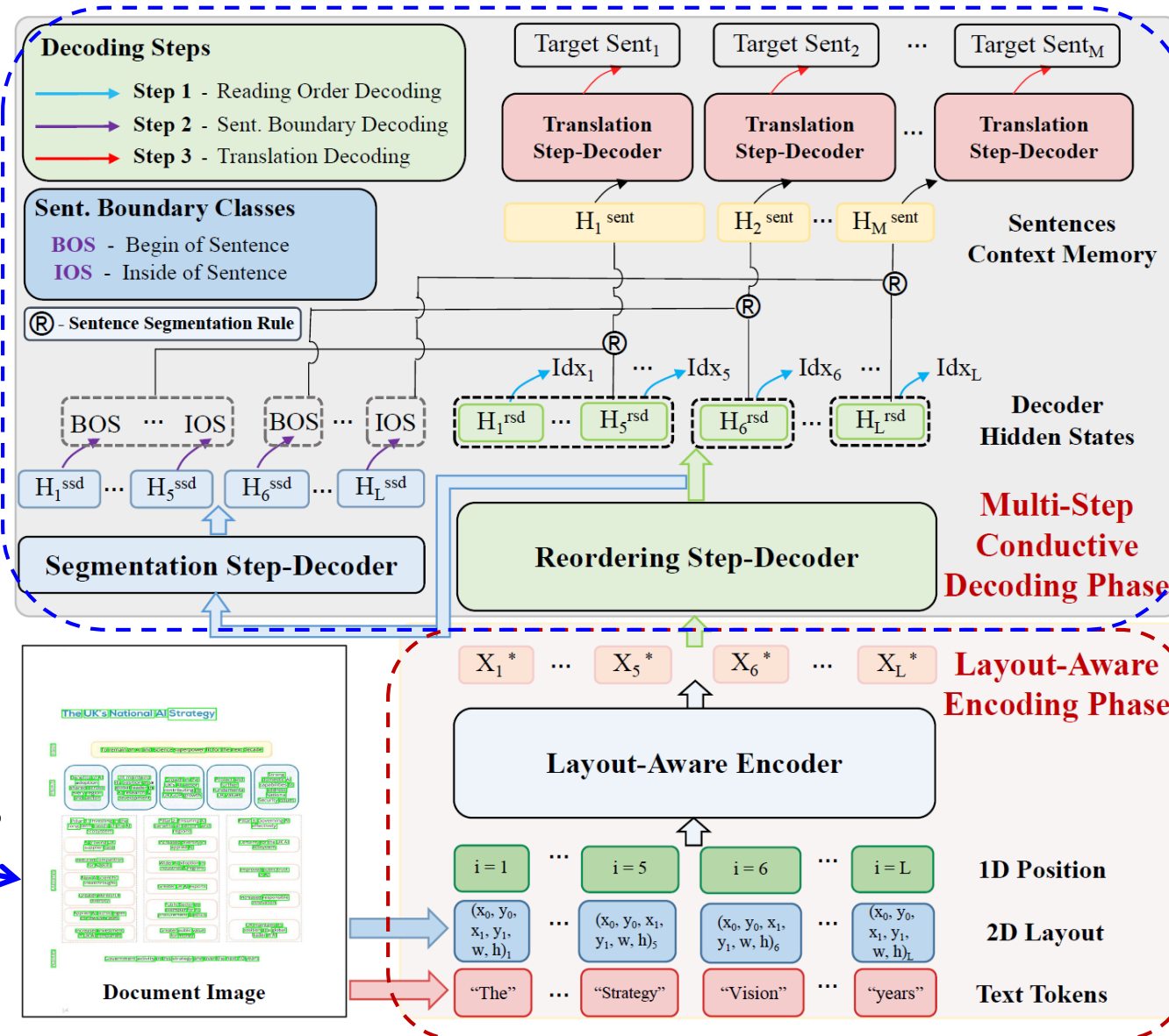
➤ **Multi-step conduction decoder (MCD)**

- The multi-step decoder gradually infers the necessary intermediate results until the final translation;
- The layout aware encoding stage and the translation decoding stage are combined end-to-end to maximize collaboration between modules and reduce error accumulation.



4. Some recent work on MT

FRAMEWORK



- ① Decoding text order and sentence boundaries to obtain logical sentence features;
- ② Based on sentence features, decode the target language translation sentence by sentence;
- ③ Feature conduction connects the encoder and various sub decoders for end-to-end joint optimization.

← MCD

← LIAE

OCRer extracts the text and layout coordinates of each word. →

- ① Encoding layout features;
- ② Integrating layout and text features at the input end;
- ③ Deep contextual fusion of layout text features.

4. Some recent work on MT

- Experiments

Table 1: Dataset Statistics

Dataset	ReadingBank	DITrans		
Domain	General	Political Report	Newspaper	Advertisement
# Page	410K	902	396	377
# Word Avg.	196	245	219	123
# Sent. Avg.	21	24	16	16
Train/Test (# Page)	400K/10K	790/112	347/49	330/47
Train/Test (# Sent.)	8,467K/211K	19,210/2,698	5,501/929	5,303/815

4. Some recent work on MT

● Experiments

Table 2: Results on the General-Domain Dataset

#	Method	End-to-End	# Params (M)	BLEU
1	DocHandler-1 (Sable et al., 2023)	-	142	30.47
2	DocHandler-2 (Sable et al., 2023)	-	172	37.75
3	LayoutLM-Dec	✓	232	45.54
4	LiLT-Dec	✓	250	45.79
5	LayoutDIT-Cascade	-	293	48.29
6	LayoutDIT	✓	206	48.20

- LayoutDIT significantly surpasses the traditional cascading method DocHandler^[2] in document translation quality in general domain scenarios.
- It is significantly superior to other end-to-end methods with layout awareness capabilities LayoutLM^[3]-Dec and LiLT-Dec^[4].

4. Some recent work on MT

● Experiments

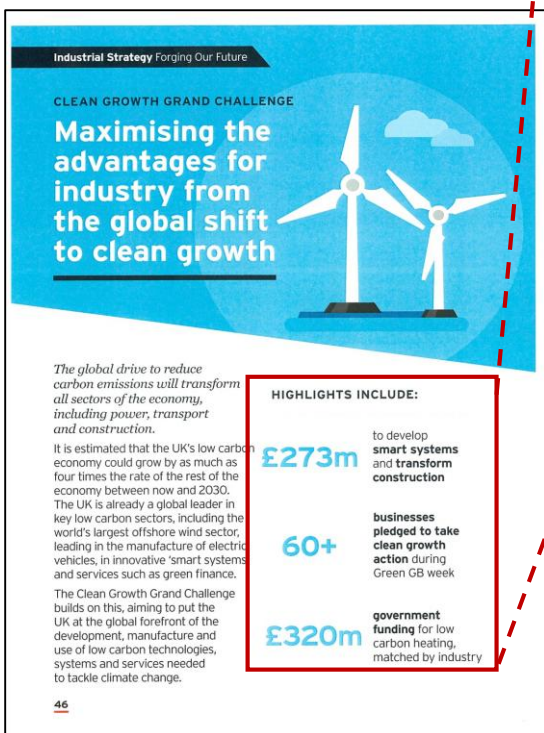
Table 3: Results in the Three Specific Domains of Our DITrans Dataset

#	Method	End-to-End	# Params (M)	Multi-Domain Learning			Zero-Shot Cross-Domain Transfer		
				Political Report	Newspaper	Advertisement	Political Report (SD)	Newspaper (TD.1)	Advertisement (TD.2)
1	DocHandler-1 (Sable et al., 2023)	-	142	21.47	26.28	23.14	21.80	16.87	13.55
2	DocHandler-2 (Sable et al., 2023)	-	172	24.80	27.34	24.99	25.19	16.88	13.60
3	MGTrans-DETR (Hinami et al., 2021)	-	212	28.30	27.61	25.76	27.34	16.26	13.09
4	MGTrans-Conv (Hinami et al., 2021)	-	238	28.99	28.84	25.63	29.51	17.78	14.07
5	LayoutLM-Dec	✓	232	45.86	40.66	36.90	45.71	27.56	21.40
6	LiLT-Dec	✓	250	46.11	40.37	36.22	46.61	23.71	18.14
7	LayoutDIT-Cascade	-	293	36.78	34.56	35.36	39.85	21.89	16.62
8	LayoutDIT	✓	206	46.97	38.82	45.72	46.70	29.01	29.46

- In the specific document domains of reports, newspapers, and advertisements, the proposed method LayoutDIT significantly outperforms the best cascading method MGTrans^[1] and end-to-end method LiLT-Dec^[4].
- Our method has good zero-shot cross-domain transfer ability and is suitable for document translation in areas with scarce samples.

4. Some recent work on MT

● Example



HIGHLIGHTS INCLUDE:

£273m

to develop smart systems and transform construction

60+

businesses pledged to take clean growth action during Green GB week

£320m

government funding for low carbon heating, matched by industry

EMNLP 23
TPAMI, accepted

Model	Source Text (S) & Translation (T)
MGTrans	(1) S: HIGHLIGHTS INCLUDE: to develop smart systems T: 亮点 包括 开发 智能 系统 ✗
	(2) S: £273m and transform construction businesses pledged to take T: 在 绿色 GB 周 期间, 273 万英镑 的 转型 和 建筑 企业 承诺 采取 60 多名 清洁 增长 行动 ✗
	(3) S: 60+ clean growth action during Green GB week T: 政府 对 低收入者 的 资助 ✗
	(4) S: £320m carbon heating, matched by industry T: 按 行业 匹配, 3.2 亿英镑 的 碳 供暖 ✗
LayoutDIT (Ours)	(1) S: HIGHLIGHTS INCLUDE: £273m T: 亮点 包括 2730 万英镑 ✗
	(2) S: to develop smart systems and transform construction T: 开发 智能 系统 和 改变 建设 ✓
	(3) S: 60+ T: 60 + ✓
	(4) S: businesses pledged to take clean growth action during Green GB week T: 企业 承诺 在 绿色 GB 周 期间 采取 清洁 的 增长 行动 ✓
	(5) S: £320m T: 3.2 亿英镑 ✓
	(6) S: government funding for low carbon heating, matched by industry T: 政府 对 低碳 加热 的 资金, 与 行业 匹配 ✓

4. Some recent work on MT

◆ Datasets

Name	Papers	Size	Brief Introduction
E2ETIMT	ICPR 2022	3,016,282	It is an end-to-end text line-level image text translation dataset, with annotation information composed of <image, source language text, target translation> triplets, including synthesis domain, subtitle domain, and street view domain, covering three translation directions: English-Chinese, English-German, and Chinese-English.
DITtrans	EMNLP 2023	1,675	It is a multiple granularity end-to-end document image translation dataset. For English-Chinese translation, the high-quality multi-granularity information has been manually annotated, such as words and their positions, sentence separation identification, sentence reading order, layout area type, layout position, and others. It can be used for tasks such as sentence reading order, discourse level OCR, and discourse level document image translation. The topics cover different fields such as government reports, news, and advertisements.
DITtrans v2	TPAMI (accepted)	411,796	It is a multi granularity end-to-end document image translation dataset that has been expanded in terms of data size, translation direction, and domain based on DITtrans. It includes general domains and four specific domains: reports, news, advertisement, and books, covering three translation directions: English-Chinese, English-German, and English-French.

4. Some recent work on MT

Name	Papers	Size	Brief Introduction
DIT700K	COLING 2025	718,000	It is an end-to-end document image text translation dataset in the field of network documents. Based on DITtransv2, it further increases the diversity of annotation data and proposes a new automatic data construction method, covering three translation directions: English-Chinese, English-German, and Chinese-English.
DoTA	NAACL 2024	126,345	It is a structured document image translation dataset for academic papers, which allows machines to output the markdown translation structure of academic papers in a structured manner. Specifically, 18,496 papers published on arXiv between 2020 and 2023 were randomly selected, covering multiple disciplines such as physics, mathematics, biology, finance, statistics, electronic engineering, economics, and computer science. Each data contains <image, translation text markdown> annotation information.



4. Some recent work on MT

ICDAR 2025 Competition on End-to-End Document Image Machine Translation Towards Complex Layouts

<https://cip-documentai.github.io/>

Schedule:

- | | |
|---|-------------------|
| • Competition website available | December 10, 2024 |
| • Training data and baseline code available | January 10, 2025 |
| • Test data release | February 20, 2025 |
| • Submission site opens | March 20, 2025 |
| • Deadline for competition submissions | April 10, 2025 |
| • Deadline for competition reports | April 20, 2025 |

4. Some recent work on MT

(2) Video-guided Machine Translation

Video-guided Machine Translation (VMT) means to translate a source language sentence (usually a subtitle) into a target language sentence with the help of a video.



4. Some recent work on MT

● The existing methods:

- Either translate the text extracted from the given video into the target language only;
- Either convert the voice in the video into text and then translate it into the target language.

They completely ignore the speaker's tone, stress, expression, gesture and other paralinguistic information, which we believe are crucial for understanding the speaker's intention and accurately interpreting his/her meanings.

4. Some recent work on MT

- Captions (source): A lot of bugs.
- Target (Chinese): 很多虫子。

bug: 故障/ 错误;
窃听器;
虫子。



Multimodal fine-grained labels

Action: run away

Sentiment: negative

Entity: people, beach, net

Emotion: none

Place: beach

Stress: 0.97, 1.02, 0.91, 1.08

Description: Two people on a beach by the sea, one of them runs away quickly after touching a fishing net, while the other one has been standing on the right side.

TriFine



4. Some recent work on MT

● Comparison between TriFine and existing VMT datasets

Dataset	Language	Domain	# Clip	Duration	# FG	Audio	Amb	Info-spec	A-S Align
How2 (2018)	En-Pt	instruction	189K	5.8s	0	✓	✗	✗	✓
VATEX (2019)	EN-Zh	caption	41K	10s	0	✓	✗	✗	✗
VISA (2022b)	En-Ja	subtitle	40K	10s	0	✗	✓	✗	✗
MSCTD (2022)	En-Zh/De	subtitle	172K	-	1	✗	✗	✗	✗
EVA (2023b)	En-Zh/Ja	subtitle	1.4M	10s	0	✗	✓	✗	✗
BigVideo (2023)	En-Zh	subtitle	3.3M*	8s	0	✓	✓	✗	✗
MAD-VMT (2024)	En-Zh	caption	193K	-	0	✗	✗	✗	✗
TriFine (Ours)	En-Zh	subtitle	2.4M	10s	7	✓	✓	✓	✓

“# FG”、“Amb”、“Info-spec”、“A-S Align” respectively represent the types of fine-grained labels, whether there is ambiguity test set, whether there is specialized feature test set, and whether speech and caption are aligned.

Each data in TriFine contains audio and subtitles aligned with audio, and each data has 7 fine-grained labels from visual and speech.

4. Some recent work on MT

Accuracy of manual evaluation

modality	Category	Accuracy	# Samples
Visual	Location	89.50%	400
	Entity	88.00%	
	Expression	86.50%	
	Action	93.25%	
	Caption	93.75%	
Audio	Audio Sentiment	79.50%	400

Boyu Guan et al., TriFine: A Large-Scale Dataset of Vision-Audio-Subtitle for Tri-Modal Machine Translation and Benchmark with Fine-Grained Annotated Tags.

In Proc. of COLING 2025

Outline

1. Introduction
2. Performance analysis of translation by LLMs
3. What we should do?
4. Some work on MT

 **5. Conclusion**

5. Conclusion

- The impact of LLMs on various research fields including MT is irreversible. It is imperative to transform thinking, embrace new technologies, and accept and adapt to the development of new technologies;
- From the perspective of NLP technology research, there are still many challenges to be solved. There is still much to be done on how to improve NLP technologies such as MT for the practical applications;
- From the perspective of foreign language learners, AI technologies can be used to help improve their abilities and efficiency. However, they must master AI technology and strive to achieve a high level of foreign language proficiency, otherwise it is difficult to avoid the danger of being replaced by AI.

谢谢!
Thanks!

