

New Trends in Machine Translation using Large Language Models: Case Examples with ChatGPT

Chenyang Lyu

Dublin City University

Dublin

Ireland

chenyang.lyu2@mail.dcu.ie

Jitao Xu

NetEase Youdao

Beijing

China

xujitaorichard@gmail.com

Longyue Wang

Tencent AI Lab

Shenzhen

China

vinnylywang@tencent.com

Abstract

Machine Translation (MT) has made significant progress in recent years using deep learning, especially after the emergence of large language models (LLMs) such as GPT-3 and ChatGPT. This brings new challenges and opportunities for MT using LLMs. In this paper, we brainstorm some interesting directions for MT using LLMs, including stylized MT, interactive MT, and Translation Memory-based MT, as well as a new evaluation paradigm using LLMs. We also discuss the privacy concerns in MT using LLMs and a basic privacy-preserving method to mitigate such risks. To illustrate the potential of our proposed directions, we present several examples for the new directions mentioned above, demonstrating the feasibility of the proposed directions and highlight the opportunities and challenges for future research in MT using LLMs.

1 Introduction

Machine Translation (MT), especially Neural Machine Translation (NMT, [Bahdanau et al., 2015](#); [Vaswani et al., 2017](#); [Castilho et al., 2017](#); [Stahlberg, 2020](#); [Kocmi et al., 2022](#)) is a fundamental task in natural language processing (NLP) that aims to automatically translate text from one language to another. Despite decades of research, MT still faces many challenges, such as dealing with idiomatic expressions, handling rare words, and maintaining coherence and fluency in the translation. Recently, the emergence of Large Language Models (LLMs), such as GPT-3 and ChatGPT ([Brown et al., 2020](#); [Chen et al., 2021](#); [Ouyang et al., 2022](#); [Wei et al., 2022](#)), has significantly advanced the state-of-the-art in MT. The zero-shot MT performance of LLMs is even on par with strong fully supervised MT systems while LLMs can be used in various scenarios beyond MT ([Wei et al., 2022](#); [Jiao et al., 2023b](#); [Wang et al., 2023](#)).

However, MT using LLMs also poses new challenges and opportunities that require new directions

and methodologies. In this paper, we brainstorm several interesting directions for MT using LLMs, including stylized MT, interactive MT, and Translation Memory (TM) based MT, as well as a potential new evaluation paradigm of translation quality using LLMs. Stylized MT ([Sennrich et al., 2016](#); [Niu and Carpuat, 2020](#)) aims to preserve the stylistic features of the source text in the translation output, such as the tone, register, formality or genre. Interactive MT ([Knowles and Koehn, 2016](#); [Santy et al., 2019](#)) aims to facilitate the collaboration and feedback between human users and MT systems, such as through chatbots or question-answering systems. TM-based MT ([Bulte and Tezcan, 2019](#); [Xu et al., 2020](#)) tends to make use of similar translations retrieved from the TM to improve the MT performance. The new evaluation paradigm using LLMs aims to leverage the power of LLMs for more accurate and efficient evaluation of MT systems from various aspects instead of only evaluating the similarity between system outputs and references.

In addition to the new directions and methodologies, we also discuss the privacy concerns in MT using LLMs and propose basic privacy-preserving methods to mitigate the risks. Privacy in LLM-based MT is becoming increasingly important, as LLMs may inadvertently reveal sensitive information in the source text or the translation output.

To preliminarily investigate the feasibility of the interesting directions mentioned above, we present several examples using ChatGPT for MT under various scenarios, demonstrating the feasibility of the directions. Our results demonstrate the potential of the potential new directions and methodologies for enhancing the quality and diversity of MT output, as well as the importance and challenges of privacy in MT using LLMs. We conclude by highlighting the opportunities and challenges for future research in MT using LLMs and suggesting potential directions for further exploration.

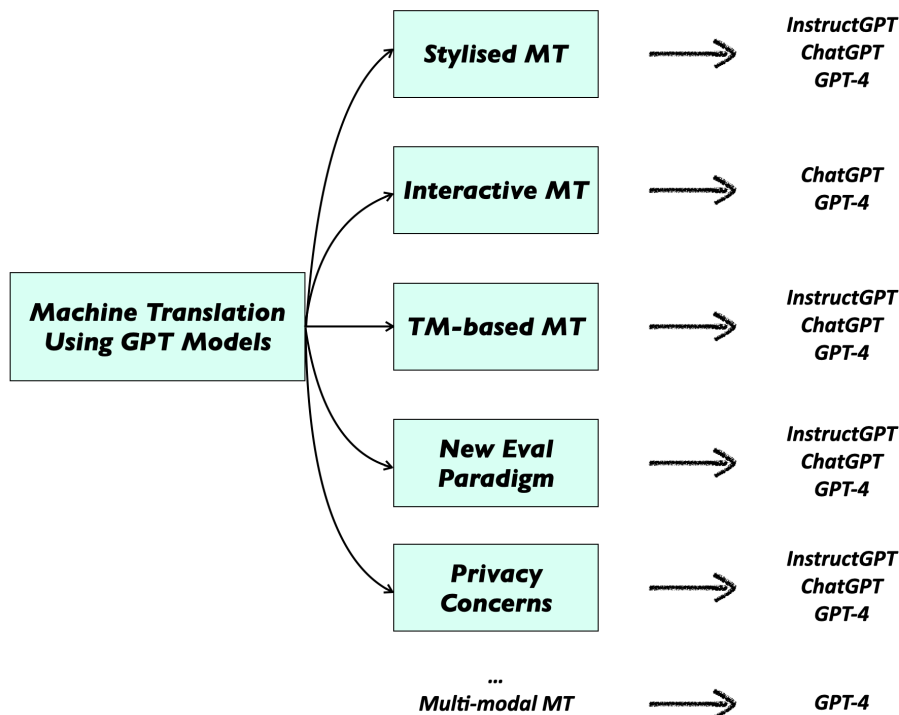


Figure 1: Interesting directions for MT using LLMs (e.g. GPT models), including stylized MT, interactive MT, Translation Memory-based MT and new evaluation paradigm for MT using LLMs, as well as privacy concerns.

2 Stylized MT

Stylized MT refers to the ability to generate translations that match a specific style or genre (Toshevskaja and Gievska, 2022; Wang et al., 2022), such as formal or informal language (Sennrich et al., 2016), poetry or prose, or different dialects or registers. This can be achieved by training MT systems on multi-parallel data that contain translations in different styles or genres, or by using style transfer techniques (Yang et al., 2018; Jin et al., 2022) that can transform a given translation into a desired style. Stylized MT has many potential applications, such as in marketing, literature, or cultural preservation.

However, stylized MT is difficult to achieve before the presence of LLMs as there lacks of such parallel corpora for stylized MT to fit various styles while the zero-shot ability of LLMs make such tasks feasible. We can directly prompt LLMs to translate text with specific style expressed by natural language or we can firstly let LLMs translate the original text and then stylize the translation output. We present an example of translating an introduction for Olympic game from Wikipedia from English to Chinese while following poetic style in Figure 2. This example shows that ChatGPT can handle translation with poetic style, which can be

Please translated the text below to Chinese following poetic style: "The Ancient Olympic Games were a religious and athletic festival that took place every four years at the sanctuary of Zeus in Olympia, Greece. Participants came from various city-states and kingdoms of Ancient Greece, competing in mainly athletic sports, but also combat sports such as wrestling and the pankration, as well as horse and chariot racing events. It is believed that during the Games, all conflicts among participating city-states were suspended until the end of the festival, which was known as the Olympic peace or truce. However, this idea is considered a modern myth as the Greeks never actually suspended their wars. Nevertheless, the truce did provide safe passage for religious pilgrims travelling to Olympia, who were protected by Zeus."

古奥林匹克，神圣嘉年华，
奥林匹亚，宙斯的圣地在。
古希腊的子民，各方参赛，
竞技体育，搏击，马车赛。
众所周知，奥运会期间，
参赛城邦，休战和平氛围环绕。
如今这一说法，被视为神话传说，
因为希腊人从未真正停战。
然而，停战为信仰的朝圣者，
提供安全通行，受宙斯庇佑着。

Figure 2: An example of prompting ChatGPT to translate text from English to Chinese using poetic style.

hardly achieved by conventional MT systems.

Nevertheless, stylized MT also poses several challenges. One challenge is how to define and measure different styles or genres in a systematic and scalable way. Another challenge is how to evaluate the quality of stylized MT, as traditional evaluation metrics may not be sufficient to capture the diversity of stylistic variations. Overcoming these challenges requires interdisciplinary collaboration between linguists, literary scholars, and

computer scientists.

3 Interactive MT

Interactive MT (Santy et al., 2019; Jiao et al., 2023a) allows users to actively participate in the translation process, either by correcting or refining automatic translations or by providing feedback on the translation quality. This can be achieved by integrating MT systems based on LLMs with interactive user interfaces, such as chatbots or on-line forums, that allow users to engage with the translation process in real time to provide feedback and more specific requirements. Interactive MT can help to improve the accuracy and fluency of translations, especially in cases where the source language is ambiguous or the domain knowledge is limited.

However, interactive MT also raises several challenges. One challenge is how to design user interfaces that are intuitive and user-friendly, yet also informative and flexible. Another challenge is how to incorporate user feedback into the translation process in a principled and effective way. Overcoming these challenges requires insights from human-computer interaction, NLP, and user experience design.

4 Translation Memory-based MT

TM has been used for decades to help translators in basic Computer-Aided Translation systems. The general process of using TM in MT is, for a sentence to be translate, to first search for similar translations in TM using, for instance, fuzzy matching techniques, then revised or edit the retrieved similar translation in order to obtain a high quality translation. TM-based MT has already been integrated into conventional NMT systems (Bulte and Tezcan, 2019; Xu et al., 2020; Cai et al., 2021). The use of retrieved similar sentence pairs (Pham et al., 2020) seems to be a natural fit to few-shot prompting techniques when performing MT using LLMs. LLMs has emerged the In-Context Learning (ICL) ability that they can learn specific tasks through task examples given in the prompt.

However, existing works so far have mostly used randomly selected translation examples as prompts and suggest that using semantically similar examples does not significantly further improve the translation performance (Vilar et al., 2022; Zhu et al., 2023). Most of these works used sentence level embedding built by an external model to retrieve

LY Please translate the text below to Chinese: "Jordan played college basketball for three seasons under coach Dean Smith with the North Carolina Tar Heels. As a freshman, he was a member of the Tar Heels' national championship team in 1982. Jordan joined the Bulls in 1984 as the third overall draft pick, and quickly emerged as a league star, entertaining crowds with his prolific scoring while gaining a reputation as one of the game's best defensive players. His leaping ability, demonstrated by performing slam dunks from the free-throw line in Slam Dunk Contests, earned him the nicknames "Air Jordan" and "His Airness". Jordan won his first NBA title with the Bulls in 1991, and followed that achievement with titles in 1992 and 1993, securing a three-peat. Jordan abruptly retired from basketball before the 1993-94 NBA season to play Minor League Baseball but returned to the Bulls in March 1995 and led them to three more championships in 1996, 1997, and 1998, as well as a then-record 72 regular season wins in the 1995-96 NBA season. He retired for the second time in January 1999 but returned for two more NBA seasons from 2001 to 2003 as a member of the Washington Wizards. During the course of his professional career he was also selected to play for the United States national team, winning four gold medals, while also being undefeated."

乔丹在北卡罗来纳大学的塔贺尔队，跟随教练迪恩·史密斯打了三个赛季的大学篮球。作为新生，他成为了1982年塔贺尔队获得全国冠军的成员。1984年，乔丹以第三顺位被芝加哥公牛队选中，并迅速成为联盟明星，他以他出色的得分和防守表现取悦了球迷们并获得了“空中飞人”和“他的空气高度”的绰号。乔丹在1991年带领公牛队获得了他的第一个NBA冠军，接着在1992年和1993年获得了冠军，完成了三连冠。乔丹在1993-1994赛季之前突然退役去打小联盟棒球，但在1995年3月回归公牛队，带领球队获得了1996年、1997年和1998年的三连冠，并创下了1995-1996赛季常规赛的72场胜利的纪录。他于1999年1月第二次退役，但从2001年到2003年又作为华盛顿奇才队的成员回归NBA打了两个赛季。在他的职业生涯中，他还入选了美国国家队，并赢得了四枚金牌，同时保持不败纪录。

(a)

LY Please translate it to Chinese while keep the people's name (such as Jordan) and organisation's name (such as Chicago Bull) in English.

Sure, here's the translation in Chinese while keeping the people's name and organization's name in English:

乔丹在北卡罗来纳大学的教练迪恩·史密斯 (Dean Smith) 带领下，在北卡罗来纳塔巴莱赫大学 (North Carolina Tar Heels) 打了三个赛季的大学篮球。作为一名新生，他曾是塔巴莱赫大学1982年全国冠军队的一员。1984年，乔丹作为第三顺位的新秀加入了公牛队，并迅速崭露头角，以他的高效得分和优秀的防守能力赢得了球迷们的欢呼声，成为了联盟的明星球员之一。他的跳跃能力，通过在扣篮大赛中从罚球线起跳扣篮的方式展示出来，使他获得了“空中飞人” (Air Jordan) 和“他的空气” (His Airness) 的绰号。乔丹在1991年与公牛队赢得了他的第一个NBA冠军，紧随其后的是1992年和1993年的冠军，三连冠已经成为了他的囊中之物。1993-94 NBA赛季前，乔丹突然退役，转投小联盟棒球比赛，但在1995年3月重返公牛队，并带领球队赢得了1996年、1997年和1998年的三个总冠军，以及1995-96 NBA赛季创下的72场常规赛胜利的纪录。他于1999年再次退役，但在2001年至2003年间以华盛顿奇才队的球员身份回归NBA赛场，打了两个赛季。在他的职业生涯中，他还被选入美国国家队，赢得了四枚金牌，同时也保持着不败的记录。

(b)

Figure 3: An example of using ChatGPT in an interactive way for MT.

similar examples via embedding similarity search. On the contrary, other studies using fuzzy match to retrieve similar translations have shown significant improvements (Moslem et al., 2023). Therefore, the conclusion about the effectiveness of using similar translations in MT using LLMs still remains unclear. Since TMs can provide useful domain and style information that can help LLMs to generate translations that better meets the translation requirement, it is a promising direction to further study how to better integrate TMs into LLMs for MT. Figure 4 illustrates an example of prompting LLM with TMs.

Previous studies on conventional TM-based MT has also shown that conventional Transformer-based NMT system already shows the ability to make use of new TMs that has never been seen by the model during training to largely improve

J Please translate the following sentence into French. Here is a similar translation you can use: En cas d'atteinte sévère ou si le patient est dans l'incapacité d'avaler les comprimés (par ex. patients alimentés par sonde), il est recommandé de débiter le traitement par une administration intraveineuse de ciprofloxacine jusqu'à ce qu'un relais par voie orale soit possible. In severe cases or if the patient is unable to take oral suspension (e.g. patients on enteral nutrition), it is recommended to commence therapy with intravenous ciprofloxacin until a switch to oral administration is possible.

J En cas de cas grave ou si le patient est incapable de prendre une suspension orale (par ex. les patients sous nutrition entérale), il est recommandé de commencer le traitement par ciprofloxacine intraveineuse jusqu'à ce qu'un passage à l'administration orale soit possible.

Figure 4: An example of prompting ChatGPT to translate text from English to French using a similar translation retrieved from TM.

domain-specific translation during inference (Xu et al., 2020, 2022). This indicates that conventional NMT systems learns to understand the relationship between a given source sentence and a similar translation and to select useful information from the given similar translation, rather than simply remember sentences seen during training. This ability is, to some extent, similar to the ICL ability of LLMs. However, to the best of our knowledge, there does not exist research works focusing on finding the relationships between these two abilities.

5 New Evaluation Paradigm for MT using LLM

Evaluating the quality of MT using LLMs is a challenging task, as existing evaluation metrics may not be sufficient to capture the full range of translation quality. In addition, existing open-access test sets may suffer from the data contamination problem as they are possibly used during the training process of LLMs. Evaluating on these test sets cannot correctly reflect the MT performance of LLMs. A new evaluation paradigm for MT using LLMs should take into account the unique characteristics of LLM-based MT, such as the ability to generate fluent but inaccurate translations or the sensitivity to domain-specific knowledge. Possible approaches to a new evaluation paradigm include using specifically-designed human evaluations (Graham et al., 2020; Ji et al., 2022) for such systems, or even directly employ LLMs to evaluate the translation output from LLMs (Kocmi and Federmann, 2023) - although studies show that LLMs would prefer the translation output from LLMs instead of other systems (Liu et al., 2023). Besides, we can also use extrinsic evaluation - we use the translation output in other tasks (Lyu et al., 2021) and measure the corresponding performance in-

LY Please translate this tweet into Chinese: "For the next 12 months, Twitter will keep none of the money."

You will receive whatever money we receive, so that's 70% for subscriptions on iOS & Android (they charge 30%) and ~92% on web (could be better, depending on payment processor).

After first year, iOS & Android fees drop to 15% and we will add a small amount on top of that, depending on volume.

We will also help promote your work. Our goal is to maximize creator prosperity.

At any point, you can leave our platform and take your work with you. Easy in, easy out."

J 在接下来的12个月中，Twitter不会留下任何收入。您将收到我们获得的所有款项，因此iOS和Android的订阅费用为70%（他们收取30%），Web为约92%（取决于支付处理器，可能会更好）。第一年后，iOS和Android的费用降至15%，我们还将根据数量增加一小笔费用。我们还将帮助推广您的作品。我们的目标是最大化创作者的繁荣。您可以随时离开我们的平台并带走您的作品。轻松进入，轻松退出。

(a)

LY How good is the translation above with some specific aspects you can think of?

J As an AI language model, my translations are not perfect, and there is always room for improvement. However, I strive to provide the most accurate translation possible. Some specific aspects to consider when evaluating the translation above include:

1. Accuracy: Does the translation convey the same meaning as the original text?
2. Grammar and syntax: Are the grammar and syntax correct, and does the translation flow naturally?
3. Cultural appropriateness: Is the translation culturally appropriate and relevant to the target audience?
4. Terminology: Does the translation use appropriate and accurate terminology in the target language?

Overall, the translation above seems to be accurate and convey the same meaning as the original text. However, it is always best to have a human translator review the translation to ensure it is appropriate and relevant for the target audience.

LY Evaluate the translation you gave above using these criteria.

J Based on the provided criteria, here's my evaluation of the translation:

Accuracy: The translation accurately conveys the meaning of the original tweet without distorting any important details.

Clarity: The translation is clear and easy to understand, using appropriate vocabulary and grammar structures.

Naturalness: The translation reads naturally and flows well, without any awkward or unnatural phrasing.

Use of appropriate Chinese language and conventions: The translation uses appropriate Chinese language and conventions, such as using simplified characters and following standard Chinese grammar rules.

Overall, the translation appears to be of good quality and should be easily understood by Chinese speakers.

(b)

Figure 5: An example of using ChatGPT for evaluating translation output.

stead directly assessing the translation quality. An example of using ChatGPT to evaluate the translation output for a tweet from Elon Musk is shown in Figure 5.

However, developing a new evaluation paradigm also poses several challenges. One challenge is how to balance the trade-off between evaluation efficiency and evaluation quality, as human evaluations can be time-consuming and expensive and

LLM-based evaluation can be biased. Another challenge is how to ensure the reliability and validity of the evaluation results, as different evaluators may have different subjective judgments or biases. Overcoming these challenges requires rigorous experimental design, statistical analysis, and transparency in reporting.

6 Privacy in MT using LLM

As LLMs become more powerful and widely used in MT, there are growing concerns about privacy and security (Xie et al., 2023). In particular, LLMs may inadvertently reveal sensitive information in the source text or the translation output, such as personally identifiable information, confidential business data, or political opinions. Privacy in MT using LLMs aims to mitigate these risks by developing privacy-preserving methods that can protect the confidentiality and integrity of the translation process.

One basic approach to privacy in MT using LLMs is to anonymize sensitive information in the textual input and then pass it to LLMs and get the output, which is then de-anonymized. An example of such issue using LLMs is shown in Figure 6. This is similar to methods integrating terminologies or user dictionary into conventional NMT systems (Crego et al., 2016).

However, privacy in MT using LLMs also poses several challenges. One challenge is how to balance the trade-off between privacy and accuracy, as privacy-preserving methods may introduce additional noise or distortion to the translation output. Another challenge is how to ensure the interoperability and compatibility of privacy-preserving methods across different languages, models, and platforms. Overcoming these challenges requires collaboration between experts in cryptography, privacy, and MT, as well as adherence to ethical and legal standards.

7 Future Directions

Personalized MT (Mirkin and Meunier, 2015; Rabinovich et al., 2017) - With the advancements in LLM-based MT, the focus can be shifted towards personalized MT. This approach can enable the provision of customized translations that are tailored to each user's preferences and needs. It can include translations that are adapted to the user's language proficiency, domain-specific terminology, or cultural references. One possible approach to



Figure 6: An example of privacy issue using ChatGPT for MT. The left one (a) is the example where the input is not anonymized, thus containing people's name information, etc. The right one is the example where the sensitive information in the input is anonymized.

provide personalized MT is to prompt LLMs with user-specific preferences or metadata, such as user search histories or social media posts - in other words, incorporating more context when translating text (Wang et al., 2017). The zero-shot ability of LLMs makes such tasks feasible, which is difficult to achieve in previous MT systems because such data is usually unavailable. However, personalized MT raises several challenges. One challenge is how to collect and store user-specific data in a privacy-preserving manner. Another challenge is how to measure the effectiveness of personalized MT, as traditional evaluation metrics may not capture the nuances of user preferences and needs. Overcoming these challenges requires careful consideration of ethical, legal, and technical issues.

Multi-modal MT (Yao and Wan, 2020; Sulubacak et al., 2020) - Another promising direction is multi-modal MT, which involves integrating visual, audio, or other non-textual information into the translation process. This approach can enhance the quality and accuracy of translations in various settings, such as image or video captioning, speech

recognition, and sign language translation. LLMs, such as GPT-4 (OpenAI, 2023), can be employed to develop models that can learn from multi-modal data and generate translations that accurately convey the meaning of the input. However, multi-modal MT poses several challenges, such as data heterogeneity, unbalanced datasets, and domain specificity. Overcoming these challenges would require developing novel algorithms that can learn from multi-modal data and generalize well across different modalities and domains.

8 Conclusion

In this paper, we discussed several interesting and promising research directions for MT under the scenario of using LLMs. We demonstrated case examples for stylized MT, interactive MT, TM-based MT, and a new evaluation paradigm for MT using LLMs, as well as preserving user-privacy in LLM-based MT. We also pointed out further directions like personalized MT and multi-modal translation. We hope this work can inspire further researches in this area of using LLMs for MT.

References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Bram Bulte and Arda Tezcan. 2019. [Neural fuzzy repair: Integrating fuzzy matches into neural machine translation](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1800–1809, Florence, Italy. Association for Computational Linguistics.
- Deng Cai, Yan Wang, Huayang Li, Wai Lam, and Lemao Liu. 2021. [Neural machine translation with monolingual translation memory](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7307–7318, Online. Association for Computational Linguistics.
- Sheila Castilho, Joss Moorkens, Federico Gaspari, Iacer Calixto, John Tinsley, and Andy Way. 2017. Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*.
- Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*.
- Josep Maria Crego, Jungi Kim, Guillaume Klein, Anabel Rebollo, Kathy Yang, Jean Senellart, Egor Akhanov, Patrice Brunelle, Aurélien Coquard, Yongchao Deng, Satoshi Enoue, Chiyo Geiss, Joshua Johanson, Ardas Khalsa, Raoum Khiari, Byeongil Ko, Catherine Kobus, Jean Lorieux, Leidiana Martins, Dang-Chuan Nguyen, Alexandra Priori, Thomas Riccardi, Natalia Segal, Christophe Servan, Cyril Tiquet, Bo Wang, Jin Yang, Dakun Zhang, Jing Zhou, and Peter Zoldan. 2016. [Systran’s pure neural machine translation systems](#). *CoRR*, abs/1610.05540.
- Yvette Graham, Christian Federmann, Maria Eskevich, and Barry Haddow. 2020. [Assessing human-parity in machine translation on the segment level](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4199–4207, Online. Association for Computational Linguistics.
- Tianbo Ji, Yvette Graham, Gareth Jones, Chenyang Lyu, and Qun Liu. 2022. [Achieving reliable human assessment of open-domain dialogue systems](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6416–6437, Dublin, Ireland. Association for Computational Linguistics.
- Wenxiang Jiao, Jen-tse Huang, Wenxuan Wang, Xing Wang, Shuming Shi, and Zhaopeng Tu. 2023a. [Parrot: Translating during chat using large language models](#). *arXiv preprint arXiv:2304.02426*.
- Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. 2023b. [Is chatgpt a good translator? a preliminary study](#). *arXiv preprint arXiv:2301.08745*.
- Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, and Rada Mihalcea. 2022. [Deep learning for text style transfer: A survey](#). *Computational Linguistics*, 48(1):155–205.
- Rebecca Knowles and Philipp Koehn. 2016. [Neural interactive translation prediction](#). In *Conferences of the Association for Machine Translation in the Americas: MT Researchers’ Track*, pages 107–120, Austin, TX, USA. The Association for Machine Translation in the Americas.
- Tom Kocmi, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Thamme Gowda, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Rebecca Knowles, Philipp Koehn, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Michal Novák, Martin

- Popel, and Maja Popović. 2022. [Findings of the 2022 conference on machine translation \(WMT22\)](#). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1–45, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Tom Kocmi and Christian Federmann. 2023. [Large language models are state-of-the-art evaluators of translation quality](#).
- Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023. Gpteval: Nlg evaluation using gpt-4 with better human alignment. *arXiv preprint arXiv:2303.16634*.
- Chenyang Lyu, Lifeng Shang, Yvette Graham, Jennifer Foster, Xin Jiang, and Qun Liu. 2021. [Improving unsupervised question answering via summarization-informed question generation](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4134–4148, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Shachar Mirkin and Jean-Luc Meunier. 2015. [Personalized machine translation: Predicting translational preferences](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2019–2025, Lisbon, Portugal. Association for Computational Linguistics.
- Yasmin Moslem, Rejwanul Haque, and Andy Way. 2023. Adaptive machine translation with large language models. *arXiv preprint arXiv:2301.13294*.
- Xing Niu and Marine Carpuat. 2020. [Controlling neural machine translation formality with synthetic supervision](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8568–8575.
- OpenAI. 2023. [Gpt-4 technical report](#).
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, et al. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*.
- Minh Quang Pham, Jitao Xu, Josep Crego, François Yvon, and Jean Senellart. 2020. [Priming neural machine translation](#). In *Proceedings of the Fifth Conference on Machine Translation*, pages 462–473, Online. Association for Computational Linguistics.
- Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. [Personalized machine translation: Preserving original author traits](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1074–1084, Valencia, Spain. Association for Computational Linguistics.
- Sebastin Santy, Sandipan Dandapat, Monojit Choudhury, and Kalika Bali. 2019. [INMT: Interactive neural machine translation prediction](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 103–108, Hong Kong, China. Association for Computational Linguistics.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. [Controlling politeness in neural machine translation via side constraints](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 35–40, San Diego, California. Association for Computational Linguistics.
- Felix Stahlberg. 2020. Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69:343–418.
- Umut Sulubacak, Ozan Caglayan, Stig-Arne Grönroos, Aku Rouhe, Desmond Elliott, Lucia Specia, and Jörg Tiedemann. 2020. Multimodal machine translation through visuals and speech. *Machine Translation*, 34:97–147.
- Martina Toshevska and Sonja Gievska. 2022. [A review of text style transfer using deep learning](#). *IEEE Transactions on Artificial Intelligence*, 3(5):669–684.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.
- David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2022. [Prompting palm for translation: Assessing strategies and performance](#).
- Longyue Wang, Chenyang Lyu, Tianbo Ji, Zhirui Zhang, Dian Yu, Shuming Shi, and Zhaopeng Tu. 2023. Document-level machine translation with large language models. *arXiv preprint arXiv:2304.02210*.
- Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu. 2017. [Exploiting cross-sentence context for neural machine translation](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2826–2831, Copenhagen, Denmark. Association for Computational Linguistics.
- Yifan Wang, Zewei Sun, Shanbo Cheng, Weiguo Zheng, and Mingxuan Wang. 2022. Controlling styles in neural machine translation with activation prompt. *arXiv preprint arXiv:2212.08909*.

- Jason Wei, Maarten Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2022. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*.
- Shangyu Xie, Wei Dai, Esha Ghosh, Sambuddha Roy, Dan Schwartz, and Kim Laine. 2023. Does prompting language model ensure privacy? *arXiv preprint arXiv:2304.03472*.
- Jitao Xu, Josep Crego, and Jean Senellart. 2020. [Boosting neural machine translation with similar translations](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1580–1590, Online. Association for Computational Linguistics.
- Jitao Xu, Josep Crego, and François Yvon. 2022. [Bilingual synchronization: Restoring translational relationships with editing operations](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8016–8030, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Zichao Yang, Zhiting Hu, Chris Dyer, Eric P Xing, and Taylor Berg-Kirkpatrick. 2018. Unsupervised text style transfer using language models as discriminators. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 7298–7309.
- Shaowei Yao and Xiaojun Wan. 2020. [Multimodal transformer for multimodal machine translation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4346–4350, Online. Association for Computational Linguistics.
- Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Lingpeng Kong, Jiajun Chen, Lei Li, and Shujian Huang. 2023. [Multilingual machine translation with large language models: Empirical results and analysis](#).